Structural and mechanistic studies of DNA repair proteins

A thesis submitted to the University of Sussex for the degree of Doctor of Philosophy

By

Peter Hornyak

September 2015

Declaration

I hereby declare that this thesis has not been and will not be submitted, in whole or in part to another University for the award of any other degree.

Peter Hornyak

Acknowledgements

Firstly, I would like to thank to my supervisors Professor Keith Caldecott and Dr Antony Oliver for allowing me to pursue a PhD at the Genome Damage and Stability Centre. Their continuous support and immense knowledge were my guiding light during my PhD.

I also would like to express my gratitude to all the members of the GDSC and my fellow lab mates from both the Caldecott and the Oliver lab for the stimulating discussions and for the nice atmosphere, which made even a hard day or a week much more enjoyable. In particular, I am grateful to Fernando, Claire, Mohan, Mat, Raquel, Gabi and Stuart for their help and guidance. Andrew also has to be acknowledged; his colourful character has been really entertaining during these years. I would also like to thank to my family for believing in me and for letting me pursue my dream.

University of Sussex

Peter Hornyak

Doctor of Philosophy Biochemistry

Structural and mechanistic studies of mammalian DNA repair proteins

Summary

Project 1: Small molecule inhibitors of TDP2

DNA Topoisomerase II (TOP2) has important roles in many cellular processes such as DNA replication and transcription, as well as in chromosome segregation. The main enzymatic function of TOP2 is to alter DNA topology and release torsional stress, by transiently introducing a double strand break (DSB) into a DNA duplex, passing a second intact duplex through the break, and then re-sealing the break. This enzymatic process involves the formation of TOP2-DNA covalent complexes, where the catalytic tyrosine (Y821) is linked to the 5' phosphate group of a substrate DNA. TOP2 'poisons' such as etoposide, doxorubicin and mitozantrone, which have found utility as anti-cancer agents, lead to an accumulation of these covalent complexes, leading eventually to cell death in rapidly replicating and dividing cells.

As many tumours treated with TOP2 poisons go on to develop chemo-resistance, it is postulated that dual-combination therapy with inhibitors of a second enzyme, 5'-tyrosyl DNA phosphodiesterase-2 (TDP2) may prevent this from occurring; TDP2 acts to remove TOP2-DNA adducts, liberating DNA ends for repair. Inhibitors of TDP2 may also prove useful as a mono-therapy in defined tumour types.

As part of an ongoing collaboration with the Sussex Drug Discovery Centre (SDDC), the aim of the project was to determine high-resolution X-ray crystal structures of TDP2 in complex with a series of deazaflavin inhibitors. The information acquired will guide

ongoing structure-based drug design, with the aim of developing and nominating a hit-to-lead compound in the near future.

Project 2: The XRCC1 phosphate-binding pocket binds poly(ADP-ribose)

In living organisms, genomic DNA is constantly exposed to both endogenous and exogenous sources of DNA damaging agents, which if not repaired, can result in the accumulation of mutations and chromosomal aberrations. Cells have evolved a series of DNA-damage repair enzymes and pathways, to cope with this perpetual threat. Poly(ADP-ribose) polymerase 1 (PARP1) is the founding member of the large ADP ribosyl transferase superfamily. Among its broad range of functions, PARP1 can detect the presence of both single- and double-strand breaks (SSBs and DSBs) in DNA, upon which it becomes catalytically activated. As a result, PARP1 then synthesises poly(ADP-ribose) polymer using NAD^+ as a co-factor, thereby modifying both itself (auto-ribosylation) and other proteins (trans-ribosylation) in the vicinity of the DNA break.

During the initial phases of the single-strand break repair (SSBR), the scaffold protein XRCC1 is recruited by PARP1, via an interaction between poly(AD-ribose) (PAR) and the central BRCT1 domain in XRCC1. However, further investigation is required to elucidate the mechanism by which the BRCT1 domain interacts with PAR. This project aims to address this question.

Contents

List of figures

List of tables

Abbreviations

α, β	3'- α, β unsaturated aldehyde
ADME	Absorption, bioavailability, metabolism and excretion
Ade	Aldehyde
ADPr	Adenosine 5'-diphosphoribose
AIF	Apoptosis Inducing Factor
AMPPNP	Adenosine 5′-(β,γ-imido)triphosphate
AOA1	Ataxia oculomotor apraxia
AP site	Apurine/Apyrimidine
APE1	Apurine/Apyrimidine endonuclease 1
APLF	Aprataxin and PNKP like factor
APTX	Aprataxin
ART	(ADP-ribosyl) transferase
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
BER	Base excision repair
bp	Base-pair
BRCA1	Breast cancer associated protein 1
BRCT	Breast cancer associated protein 1 C-terminal domain
BSA	Bovine serum albumin
CD	Catalytic domain
C. elegans	*Caenorhabditis elegans*
CHK2	Checkpoint kinase 2
CHO	Chinese hamster ovary
CK2	Casein kinase 2
CPT	Camptothecin
C-terminus	Carboxyl terminus
CtIP	CTBP interacting protein
DDR	DNA damage response
ddTTP	Dideoxythymidine triphosphate
DMSO	Dimethyl sulfoxide

DNA	Deoxyribonucleic acid
DNA PK	DNA dependent protein kinase
DNA PKcs	DNA dependent protein kinase catalytic subunit
dNTP	Deoxynucleotide triphosphate
dRP	5'-deoxyribose phosphate
DSB	Double strand break
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
e-MART	Eukaryotic mono-ADP-ribosyltransferase
FCS	Foetal calf serum
FEN1	Flap endonuclease 1
FHA	Fork-Head Associated Domain
H2A	Histone H2A
H2AX	Histone H2A variant X
γH2AX	Histone H2A variant X phosphorylated on S139
H2B	Histone H2B
H3	Histone H3
H4	Histone H4
H_2O_2	Hydrogen peroxide
HCl	Hydrogen chloride
HR	Homologous recombination
HTS	High-throughput screen
IMAC	Immobilised metal affinity chromatography
ION-X	Ion-exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kan	Kanamycin
Kb	Kilobase
KOH	Potassium hydroxide
LB	Luria-Bertani medium
Lig1	DNA ligase 1

MACROD1	Macrodomain-containing protein 1
MACROD2	Macrodomain-containing protein 2
MAR	Mono(ADP-ribose)
MDC1	Mediator of DNA damage checkpoint protein 1
MEF	Mouse embryonic fibroblast
$MgCl_2$	Magnesium chloride
min	minute
Mre11	Meiotic recombination 11
MRN	Mre11/Rad50/Nba1
NAD^+	Nicotinamide adenine dinucleotide
NBS1	Nijmegen breakage syndrome 1
NLS	Nuclear localization signal
NPPP	4-nitrophenyl phenylphosphonate
N-terminus	Amino-terminus
NTR	N-terminal region
OD	Optical density
OH	Hydroxyl
ORF	Open reading frame
p53	Tumour protein 53
PAGE	Polyacrylamide gel electrophoresis
PAR	Poly(ADP-ribose)
PARG	Poly(ADP-ribose) glycohydrolase
PARP	Poly(ADP-ribose) polymerase
PBM	PAR binding motif
PBS	Phosphate buffered saline
PBZ	PAR binding zinc-finger
PCNA	Proliferating cell nuclear antigen
PG	Phosphoglycolate
pI	Isoelectric point
PNK	Polynucleotide kinase phosphatase
Polβ	DNA polymerase beta

Polδ/ε	DNA polymerase delta/epsilon
RAD50	Radiation sensitive 50
RPA	Replication protein A
ROS	Reactive oxygen species
RNAP	RNA polymerase
RNF8	Ring finger protein 8
RNF168	Ring finger protein 168
rpm	Revolutions per minute
SAR	Structure activity relationship
SAXS	Small-angle X-ray scattering
SCAN1	Spinocerebellar axonal neuropathy 1
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SIRT	Sirtuin
S-phase	Synthesis phase
SS	Single-strand
SSB	Single-strand break
SSBR	Single-strand break repair
TAE	Tris base, Acetic acid, EDTA
TARG1	Terminal ADP–ribose protein glycohydrolase 1
TBE	Tris base, Boric acid, EDTA
TBST	Tris-buffered saline and Tween 20
TDP1	Tyrosyl DNA phosphodiesterase 1
TDP2	Tyrosyl DNA phosphodiesterase 2
TEMED	Tetramethylethylenediamine
TMAO	Trimethylamine *N*-oxide
TOP1	DNA Topoisomerase 1
TOP2	DNA Topoisomerase 2
UBA	Ubiquitin-associated
UV	Ultraviolet
V(D)J	Variable (Diverse) Joining recombination

VPg	Viral protein genome-linked
v/v	Volume per volume
WRN	Werner
WT	Wild type
w/v	Weight per volume
XLF	XRCC4-like factor
XRCC1	X-ray cross complementing 1
XRCC4	X-ray cross complementing 4

CHAPTER ONE

Introduction

1.1 Project 1 – Small molecule inhibitors of TDP2

1.1.1 The DNA damage response (DDR)

DNA double-strand breaks (DSBs) can either emerge under pathological circumstances, from exposure to environmental agents such as ionizing radiation, or from endogenous sources of DNA damage such as oxygen free radicals or abortive activity of topoisomerases. These breaks, if left unrepaired, can lead to chromosomal rearrangements, carcinogenesis, or cell death (Richardson and Jasin 2000), (Bennett et al. 1993).

Living organisms evolved a delicate repair system to maintain their genetic integrity. When DSBs occur, cells activate a signal transduction network, termed the DNA damage response (DDR) (Jackson and Bartek 2009). DDR has multiple functions: controlling the cell cycle, remodelling the chromatin to allow the repair system to access the break site and recruiting DNA repair proteins to mend the lesions. To undertake these tasks, DDR employs a multitude of repair proteins with different roles. First, DNA lesions are recognised by sensor proteins, such as the MRN complex (composed of Mre11, Rad50 and Nbs1 proteins) or the Ku complex. These sensor molecules then recruit downstream transducer molecules, e.g. ataxia telangiectasia mutated (ATM) protein to amplify the signal by phosphorylating various target molecules (Uziel et al. 2003) At DSB sites, dimeric ATM autophosphorylates itself and dissociates into active monomers (Bakkenist and Kastan 2003). The activated ATM then rapidly phosphorylates H2Ax, a histone H2 variant, at serine 139 (γ-H2Ax) (Rogakou et al. 1998). This modification extends approximately one megabase in both directions on DNA and serves as a platform to recruit further factors. ATM also phosphorylates checkpoint kinase 1 and 2 (Chk1 and Chk2) resulting in delay in cell cycle (Matsuoka et al. 1998). The modification of SMC1 by ATM leads to chromatin remodelling to provide an open platform for DNA repair protein (Yazdi et al. 2002). Depending on the severity of the damage, cells either attempt to repair the DNA lesion by arresting or delaying the cell cycle or commence apoptosis if the DNA damage is too extensive by phosphorylating P53 by ATM which can transactivate its targets NOXA, BAX and PUMA (Banin et al. 1998) (Yu and Zhang 2008).

1.1.1.1 Recognition and repair of chemically diverse DNA ends

DNA strand breaks rarely possess ‘clean’ 3’-hydroxyl and 5’-phosphate moieties that are ready for ligation. Instead, the DNA repair machinery deals with chemically divergent

DNA termini by employing a wide variety of end processing enzymes. Damaged DNA termini can result from ionizing radiation, reactive oxygen species (ROS), abortive ligation events, and topoisomerase-induced protein-DNA crosslinks. Radiation- and ROS-induced base damage and DNA strand breaks are estimated to arise at a frequency of 10,000 per cell per day and generate DNA ends harbouring a fragmented sugar, 3'-phosphate, or 5'-hydroxyl termini (Lindahl and Nyberg 1972) (Ames et al. 1993). Functional impairment of the end processing factors can result in severe neuronal diseases (Rulten and Caldecott 2013). Polynucleotide kinase 3'-phosphatase (PNKP) is able to restore 3'-phosphate and 5'-hydroxyl termini to their canonical chemistry, which is 3'-hydroxyl and 5'-phosphate, and loss of PNKP activity in humans results in *microcephaly with early onset seizures* (MCSZ) and *ataxia with oculomotor apraxia type 4* (AOA4) (Bras et al. 2015). Abortive ligation events can result in DNA adenylation and the formation of breaks with 5'-AMP ends, which are resolved by Aprataxin (APTX); mutation of which result in *ataxia with oculomotor apraxia 1* (AOA1)(Moreira et al. 2001). Abortive activity of topoisomerases can generate DNA SSBs and DSBs with 5'- or 3'-termini that are covalently linked to the topoisomerase, and these adducts are processed by the tyrosyl DNA phosphodiesterases, TDP1 and TDP2 (Yang et al. 1996, Cortes Ledesma et al. 2009). Mutations in TDP1 result in *spinocerebellar ataxia with axonal neuropathy-1* (SCAN1), whereas mutations in TDP2 are associated with seizures and ataxia (El-Khamisy et al. 2005, Gomez-Herreros, Schuurs-Hoeijmakers et al. 2014).

Damaged DNA termini can also be processed by endo- and exonucleases. For example, the exonucleases FEN1 and the MRN/CtIP complex resolve 5'-end blocking lesions (Murray et al. 1994, Chen et al. 2008). Artemis nuclease has several different nucleolytic activities; its 5'-endonucleoltytic activity nicks 5' overhangs to create blunt ended DNA, and its 5' to 3' exonuclease activity can act on single stranded DNA (Ma et al. 2002). Werner (WRN) protein, which is mutated in Werner syndrome, possess both helicase and 3' to 5' exonuclease activity (Huang et al. 1998).

1.1.1.2 Double strand break repair

In mammalian somatic cells the two most common repair systems are nonhomologous end joining (NHEJ) and homologous recombination-mediated repair (HR) (Figure 1.1.1) (Hartlerode and Scully 2009). NHEJ is the most frequently used repair process in non-dividing haploid organisms and in diploid organisms that are not in S phase (Lieber 2010).

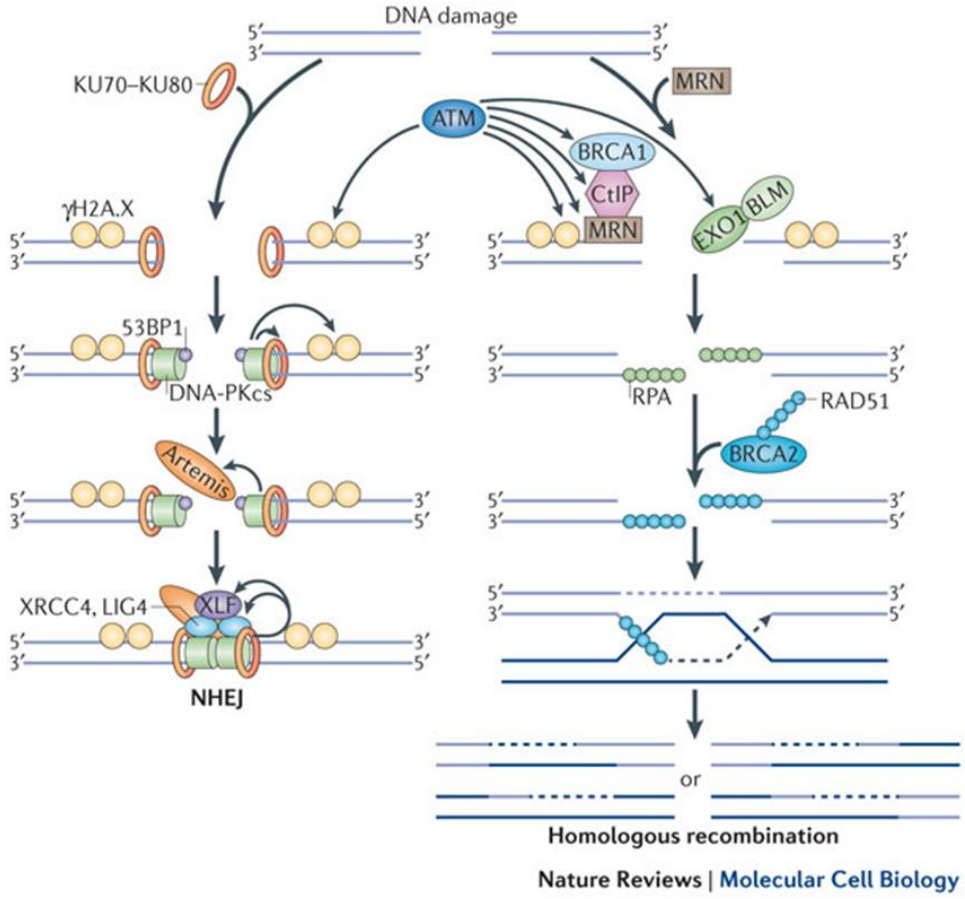

Figure 1.1.1. Schematic representation of nonhomologous end joining (NHEJ) and homologous recombination (HR). NEHJ is mainly taking place in non-proliferative cell or proliferative cells in G1 phase. DNA double strand break (DSB) are recognised and tethered by the Ku70/80 complex, which in turn recruits the catalytic domain of the DNA-dependent protein kinase (DNA-PKcs) and 53BP1 and other main NHEJ factors. DNA-PKcs phosphorylates mainly itself and several other repair proteins. When required, end processing factors restoring the damaged DNA end to their canonical chemistry, such as Artemis (depicted). Ligatable ends are joined by Lig4 in complex with XRCC4. XLF stimulates Lig4 activity. HR is preferred in replicating cells as DBSB can be repaired error free due to the presence of a sister chromatid following replication. The MRN complex recruits CtIP allowing end resection. Exposed single stranded DNA is coated by replicating protein A (RPA) which subsequently replaced by Rad51 filaments. Rad51 mediates the strand invasion and homology search on the sister chromatid which is the extended by DNA polymerases and finally ligated. Figure is adapted from (Chowdhury et al. 2013).

1.1.1.2.1 Nonhomologous End Joining (NHEJ)

Apart from its repair function NHEJ is also involved in V(D)J recombination and Class Switch Recombination CSR during the maturation of T and B cells and thus patients with defects in NHEJ are frequently immunodeficient (Lieber 2010).

During NHEJ the broken DNA ends are re-joined without the need of a repair template. Consequently, the process is potentially error prone since genetic information may be lost at the site of the break. The repair process is initiated by the binding of the Ku heterodimer (Ku70 and Ku80) to the DNA ends. The Ku-DNA complex acts as a platform for further actions performed by various proteins. Ku70/80 recruit DNA-PKcs (DNA-dependent protein kinase catalytic subunit) and its binding promotes the juxtaposition of the two DNA ends (Gottlieb and Jackson 1993). If the DSB does not require end processing, NHEJ can proceed to the ligation step where the DNA break is sealed by a ligation complex comprised of XRCC4, DNA ligase IV, and XLF (XRCC4-like factor; aka Cernunnos) (Grawunder et al. 1997) (Ahnesorg et al. 2006).

If the DNA termini require end processing or the removal of damaged nucleotides, NHEJ is able to employ nucleases for end processing as described above. In addition, DNA Polymerases Pol μ and Pol λ are both able to bind to Ku:DNA complexes via their BRCT domains to conduct DNA gap filling (Mahajan et al. 2002). Once both ends are restored, DNA ligation can occur (Lieber 2010).

1.1.1.2.2 Homologous Recombination (HR)

During S and G2 phases homology directed repair is favoured as sister chromatids stay in close proximity providing a repair template. When a DSB occurs on one of the sister chromatids the broken DNA is resected at the 5' terminus to create a 3'-single-stranded tail (Rupnik et al. 2010). The single stranded tail enables formation of a nucleoprotein filament containing the recombinase RAD51, which enables strand invasion into the homologous region of the intact sister chromatid, with the original displaced DNA strand forming a D -loop (Chen et al. 2008). The lost genetic information at the damaged site is replaced by a DNA polymerase using the intact sister chromatid as the template. There are two alternative ways to complete the repair process. In the course of DSBR (double Holliday junction model) the second 3' overhang end from the broken DNA also engages with the sister chromatid forming a second Holliday junction. Resolution of this junction by

resolvases can result in crossover or non-crossover recombinant products (Heyer 2004). The alternative route is the synthesis-dependent strand annealing (SDSA) where the synthetized strand is exchanged between the sister chromatid template and other 3' end of the repaired break (Wyman and Kanaar 2006).

1.1.2 Topoisomerase II (TOP2)

1.1.2.1 Topoisomerase function and families

Topoisomerases have indispensable roles in the organization and maintenance of the integrity of the chromosomal DNA. By introducing transient breaks topoisomerases can release torsional stress, decatenate chromosomes and change the level of supercoiling of DNA (Interthal et al. 2001). As transcription and replication machineries slide along the DNA duplex they generate positive supercoiling, which tightens the DNA duplex and prevents further strand opening (Wu et al. 1988). Negative supercoiling emerges behind the acting polymerases, which can lead to non-canonical DNA structures such as R-loops, hairpins or guanosine quartets (Aguilera and García-Muse 2012).Topoisomerases are found in all kingdoms of life and are of two main classes: type I and type II. Type I enzymes cleave one strand of the DNA, and type II topoisomerases cut both strands. A common feature of all type of topoisomerases is that they break the DNA by transesterification; an active site tyrosine performs a nucleophile attack on the DNA phosphodiester backbone and becomes covalently attached to the DNA during the process (Krogh and Shuman 2000).

Topoisomerases type IA and type IB, which are subclasses of the type I enzyme family, are able to release supercoiling by cutting one strand on the DNA. Type IA facilitates the passage of an intact single DNA strand through the opened strand to release torsional stress. In contrast, the type IB enzyme allows the rotation of the DNA duplex around the intact phosphodiester bond. The reaction is driven by the energy of the DNA supercoiling (Corbett and Berger 2004). The human Topoisomerase 1 (TOP1) is a type IB topoisomerase and becomes covalently attached to the 3' end of the cleaved DNA. Type IIA and IIB topoisomerases (subclasses of the type II family) facilitate the passing of an intact duplex through a transient DNA double strand break. Eukaryotic type IIA topoisomerases are homodimeric enzymes and in order to catalyse DNA cleavage and strand passage they require divalent metallic ions such as Mg^{2+} and the binding and hydrolysis of ATP (Sissi and Palumbo 2009).

1.1.2.2 TOP2 enzymes

Mammalian cells express two type IIA isoforms: Topoisomerase IIα and IIβ (TOP2α and TOP2β) (Chung et al. 1989). The two isoforms show similar structural features and catalytic activity. TOP2α is essential for all dividing cells and indispensable in the separation of sister chromatids and its expression is increased from S phase to M during the cell cycle (Kimura et al. 1994). TOP2β is required for neuronal development and transcriptional regulation but is dispensable in some cell types (Yang et al. 2000, Lyu et al. 2006).

1.1.2.2.1 Enzymatic mechanism of TOP2

The enzyme mechanism of topoisomerase type II from yeast (*S. cerevisiae*) was studied extensively by biochemical and structural methods. As previously mentioned, TOP2 is homodimeric and each subunit breaks and transiently binds one DNA strand via a phosphotyrosyl bond (Dong and Berger 2007). The domain organization of each subunit can be seen in Figure 1.1.2. The N-terminal domain is an ATP binding domain followed by the TOPRIM domain, which is a well-conserved structure found not only in topoisomerases but also primases and other DNA binding enzymes (Aravind et al. 1998, Classen et al. 2003). The TOPRIM domain is involved in cation binding required for the catalytic step. The central part contains the winged helix domain (WHD) harbouring the catalytic tyrosine residue (Y782) and the Tower (or shoulder) domain involved in DNA interaction (Berger et al. 1996). The C-gate is linked to the tower domain via a coil-coiled structure and it is the exit point for the intact DNA duplex. The C-terminal part is not conserved between different topoisomerase type II enzymes from different species and it has a role in nuclear localization and protein-protein interaction (Nitiss 2009) (Schmidt et al. 2012).

During decatenation or the relaxation of supercoiled DNA, TOP2 interacts with two DNA duplexes (Figure 1.1.3). TOP2 introduces a transient double strand break in one duplex termed the G-segment and helps the translocation of the second intact DNA duplex termed the T-segment (Rybenkov et al. 1997, Fass et al. 1999). The energy released by cleavage of the phosphodiester bond is retained in the newly formed phosphotyrosine bond, and so the cleavage/ligation reaction does not require ATP. However, ATP hydrolysis enables conformational changes in the enzyme that assist in T-segment passage and release of the G-segment (Schmidt et al. 2012).

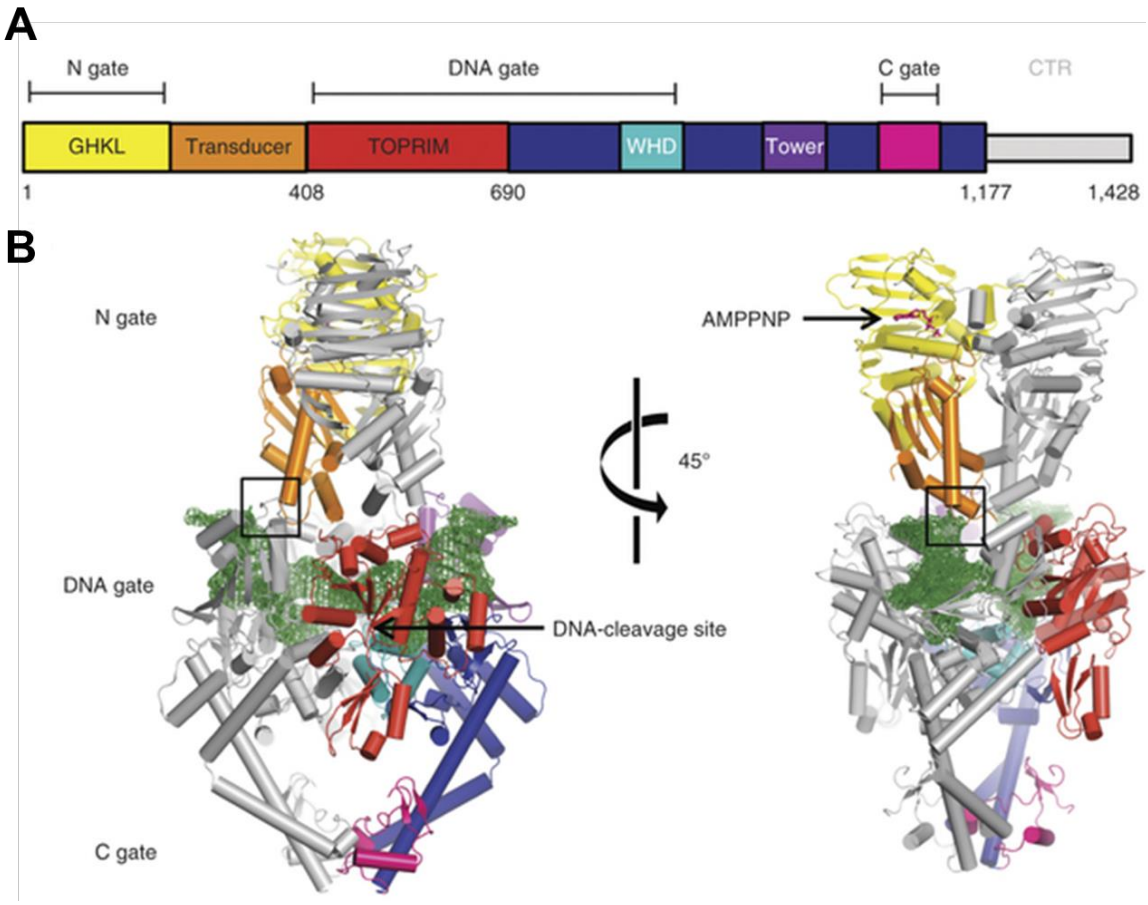

Figure 1.1.2. Structure of TOP2α binding to DNA. **A.** Schematic representation of TOP2 domain organization. **B.** Structural model of the DNA bound TOP2α ternary complex. The N-gate (in yellow) binds ATP and it is shown by AMPPNP a non-hydrolysable analogue of ATP. The TOPRIM domain (in red) responsible for cation binding. The winged helix domain (WHD, in light blue) with the catalytic Y782 residue and the Tower domain (in aubergine) are involved in DNA interaction. The Coiled-coil structure forms the C-gate (in pink). The C-terminal part is responsible for protein-protein interactions. DNA is depicted as $2F_o$-F_c density (1.5σ contour) in green. (Figure from Schmidt et al. 2012)

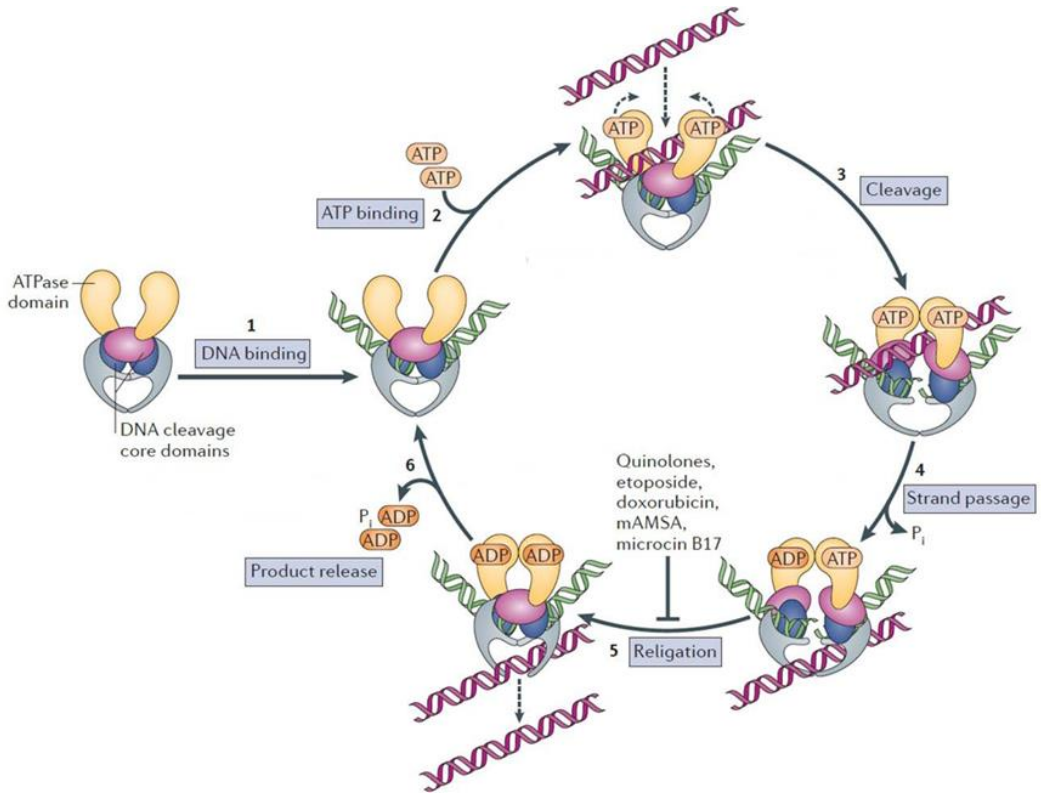

Figure 1.1.3. Type II topoisomerase reaction cycle. First, TOP2 binds the first DNA segment (termed G-segment), then by associating with the second DNA segment (T-segment) introduces a transient double strand break on the G-segment. The DNA ends are protected by a 5' phosphotyrosyl covalent bond formed between the catalytic tyrosine residue (Y782 in yeast) and the 5'phosphate end of the DNA. Following the opening of the G-segment, the topoisomerase then facilitates the passage of the T-segment. The open strands of the g-segment are religated and released. TOP2 poisons are able to prevent relegation creating TOP2-DNA adduct. Figure is adapted from Vos et al. 2011.

During strand passage each 5'-phosphate terminus of the DSB is covalently linked to the catalytic tyrosine (Y782) residue via a phosphotyrosyl bond and so is protected from processing by other enzymes. The presence of these transient breaks (denoted cleavage complexes) makes TOP2 a potentially dangerous protein, however. Whilst cleavage complexes are obligate intermediates in topoisomerase activity, on occasion they can become abortive: the TOP2-DNA complex is stabilised and prevented from religating the DNA. These abortive covalent complexes can emerge naturally, due to oxidative stress, acidic pH and DNA modification (e.g. abasic site), and can be converted into permanent DSBs by collision with replication forks or transcription complexes. Another source of trapped topoisomerase-DNA complexes is via exposure to TOP2 poisons (see sections below) (Robinson and Osheroff 1990, Wilstermann and Osheroff 2003).

1.1.2.2.2 DSBs induced by abortive TOP2-DNA complexes are repaired by DSBR.

The removal of the abortive TOP2-DNA adducts is vital for cell viability (Nitiss 2009). The covalently trapped TOP2 enzyme is first targeted to proteolytic degradation by the 26S proteasome and can be repaired in several ways (Figure 1.1.4). The nucleolytic pathway employs the exonuclease complex MRN/CtIP to remove the covalent adduct (Hartsuiker et al. 2009). In resting cells DSBs are restored by the NHEJ pathway. As exonucleases remove nucleotides during DNA end resection, the process results in information loss and error prone repair. An error free, non-nucleolytic pathway involves the breakage of the phosphotyrosyl bond between the DNA and the topoisomerase. It has been known that tyrosyl DNA phosphodiesterase 1 (TDP1) catalyses the removal of the trapped topoisomerase I (TOP1) by hydrolysis of the 3'-phosphotyrosyl bond (Yang, Burgin et al. 1996). It was recently discovered that there is an enzyme that processes 5'-phosphotyrosyl bonds, denoted tyrosyl DNA phosphodiesterase 2 (TDP2). TDP2 is involved in many cellular processes such as transcription regulation and signalling pathways, and it was demonstrated that it catalyses hydrolysis of the 5'-phosphotyrosyl bond that links TOP2 to the 5'-terminus of DNA (Cortes Ledesma et al. 2009). TDP2 will be discussed in more detail in the next section. The proteolytic pathway involves the degradation of TOP2 by triggering the ubiquitin/26S proteasome pathway (Mao et al. 2001, Zhang et al. 2006, Gao et al. 2014). The remaining small peptide bound to the DNA is removed by TDP2. Both the proteolytic and the nucleolytic pathways result in DSBs that can be processed by either nonhomologous end joining (NHEJ) or homologous recombination (HR).

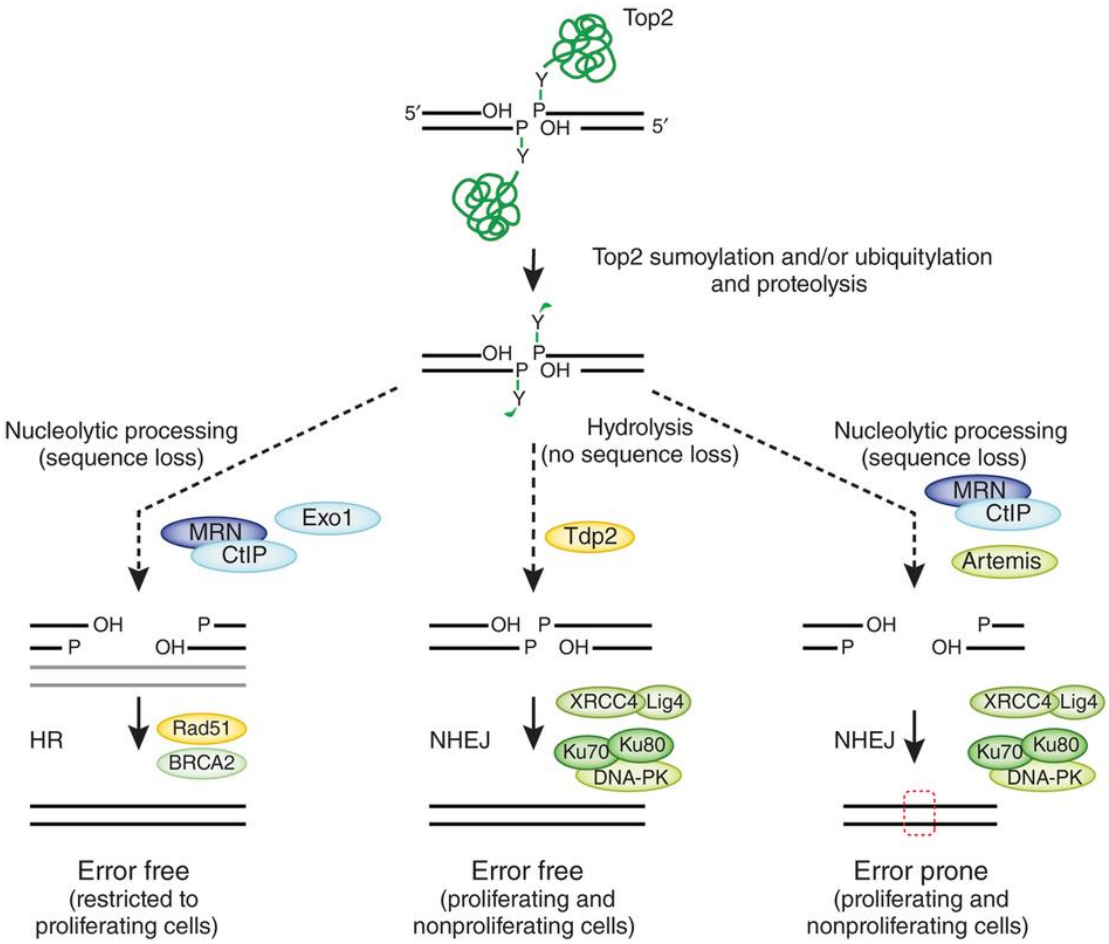

Figure 1.1.4. TDP2 is part of the NHEJ pathway. TOP2-DNA abortive complexes can either emerge naturally or from poisoning by chemotherapeutic agents. First, the complex is targeted for proteolytic degradation. Depending on the stage in the cell cycle, double strand breaks can be repaired by either NHEJ or HR. In NHEJ the nucleolytic processing of these breaks results in sequence lost and error prone DNA repair. TDP2 hydrolyses the 5' phosphotyrosyl bond without sequence lost. The repair is error free. Figure is adapted from (Caldecott, 2012).

1.1.2.2.3 TOP2 as chemotherapeutic drug target

Topoisomerase poisoning can emerge not only naturally but also from the effect of drugs (Pommier et al. 1985). Because TOP2 is required in proliferating cells, it is a useful target for inhibition in cancer chemotherapy (Nitiss 2009). TOP2 inhibitors can be categorized into two major groups according to their action on TOP2. Drugs such as novobiocin and merbarone can act as catalytic inhibitors and do not increase the level of TOP2-DNA complexes. In contrast, TOP2 poisons such as etoposide and doxorubicin can trap the enzyme intermediate covalently bound to DNA (Chen et al. 1984) . The increased level of TOP2 covalent complexes can block transcription and replication and can lead to apoptosis. These poisons were first thought to be effective in highly proliferative tumours with high expression level of TOP2α.

Etoposide was clinically developed and is widely used in a broad range of solid tumours and was approved by the FDA in 1983. The etoposide induced stabilization of TOP2-DNA complexes was studied extensively (van Maanen et al. 1988). As opposed to other TOP2-DNA complex poisoning drugs, etoposide is not an intercalating agent and can trap the TOP2 homodimers asymmetrically, resulting in mainly single stranded break formation (Chen et al. 1984) (Long et al. 1985). Etoposide also very effectively binds both TOP2α and TOP2β, whereas other drugs, such as doxorubicin are more selective towards TOP2α. Trapping TOP2β by etoposide is mainly the responsible for translocations and the induction of secondary leukaemia following chemotherapy (Cowell and Austin 2012).

TOP2α has been proven to be an effective clinical target in chemotherapy; however, recent findings suggest that TOP2β might also become a good drug target on its own right (Jarvinen and Liu 2006) (Nitiss 2009). TOP2β is recruited with androgen and oestrogen receptors to regulatory sites of hormone activated genes. TOP2β initiates transcription by introducing a transient double strand break, supposedly to resolve topological constraints during the induction of transcription (Ju et al. 2006). The break is recognised by the DNA repair machinery and repair factors such as ATM, DNA-PK and PARP1 are recruited to the damage site. Many slow proliferative cancer types (prostate and breast) can respond to hormone therapy by reinitiating their transcriptional program, therefore, it was hypothesised that TOP2β inhibition alongside with cycling hormone therapy could be an effective treatment (Haffner et al. 2011).

1.1.3 Tyrosyl DNA phosphodiesterase 2 (TDP2)

1.1.3.1 Introduction

TDP2 has diverse roles in cellular pathways and the investigation of these functions by independent studies resulted in alternative names for this enzyme such as EAPII, TTRAP and TDP2. The protein was first identified as a novel intracellular binding partner for CD40 and was termed TTRAP (TRAF and TNF receptor associated protein) (Pype et al. 2000). Yeast two-hybrid screen has shown that this protein also interacts with the transcription factor ETS1 hence the alternative designation EAPII (ETS1-associated protein II) (Pei et al. 2003). And finally, it was only recently discovered that this protein catalyses the removal of protein adducts (TOP2) from DNA by cutting the 5'-phosphotyrosyl bond (Tyrosyl DNA phosphodiesterase 2, TDP2) (Cortes Ledesma et al. 2009).

TDP2 is conserved from worms to mammals. In human this 362-amino acid protein is ubiquitously expressed in most tissues such as lung, ovary, prostate and kidney (Li et al. 2011). The protein has numerous interacting partners involved in many biochemical processes. By interacting with members of the TNF receptor superfamily and TRAFs (TNF receptor-associated factor) TDP2 negatively regulates TNFα signalling and therefore suppresses inflammatory response (Pype et al. 2000). TDP2 can interact with several transcription factors such as members of the Ets (E-twenty six) transcription factor family. Ets transcription factors control cellular proliferation, differentiation and lymphocyte development (Pei et al. 2003). By repressing the transcriptional activity of ETS1, TDP2 can negatively control the transcription of the MMP1 (matrix metalloproteinase-1) promoter. TDP2 also modulates JNK/p38 apoptosis pathway in neuronal cells by interacting with DJ-1. Mutated DJ-1 is one of the several genetic factors that are responsible for the autosomal recessive Parkinson disease (Li et al. 2011). It was also demonstrated that TDP2 is associated with virus response (Lee et al. 2003).

1.1.3.2 TDP2 hydrolyses 5'-tyrosyl DNA phosphodiester bonds

TDP2, previously known as TTRAP, was identified in a screen for new tyrosyl DNA phosphodiesterase activities (Cortes Ledesma et al. 2009). *Tdp1*/*Rad1* deleted *S. cerevisiae* yeast cells were transformed with a human complementary DNA library and were screened for resistance to the TOP1 poison camptothecin (CPT). Compared to TDP1, which liberates abortive TOP1-DNA complexes by cleaving the 3'-tyrosyl DNA phosphodiester bond, recombinant TDP2 showed ~50-fold lower 3'-tyrosyl DNA phosphodiesterase

activity *in vitro*. However, the enzyme exhibited robust 5'-tyrosyl DNA phosphodiesterase activity.

TDP2 is the primary source of 5'-tyrosyl DNA phosphodiesterase activity in vertebrate cells (Zeng et al. 2011). In TDP1 deleted vertebrate cells, TDP2 can also contribute to 3' tyrosyl DNA phosphodiesterase activity, albeit weakly, since DT40 cells in which both TDP1 and TDP2 are deleted are more sensitive to the TOP1 poison camptothecin than cells in which only TDP1 is deleted (Zeng et al. 2012)

1.1.3.2 TDP2 -dependent NHEJ protects against genome instability

TDP2 is the major, if not only, 5'-tyrosyl DNA phosphodiesterase in mammals (Gomez-Herreros et al. 2013). TDP2 knockout mice do not exhibit any obvious abnormal pathology, suggesting that TDP2-independent pathways are able to repair most endogenous TOP2 damage in the absence of TDP2. However, these mice have shown extreme sensitivity to TOP2 poisons. Etoposide administration results in increased mortality due to intestinal damage and toxicity in lymphoid tissues. It was also shown that TDP2 is a component of the NHEJ pathway as it was demonstrated that Ku70 and TDP2 exhibit an epistatic relationship. TDP2 loss results in an elevated DSB repair defect in the cell cycle stages G0/G1 and G2 (Gomez-Herreros et al. 2013).

1.1.3.3 TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function

The removal of topoisomerase-DNA adducts is crucial to protect the genetic integrity not only in cells, but also at the level of the whole organism. Impeded DNA phosphodiesterase function can result in neurodegenerative disorders. SCAN1 (*spinocerebellar ataxia with axonal neuropathy-1*) is a hereditary human disease, affecting terminally differentiated neurons (Takashima et al. 2002). In this disease, a recessive mutation in the active site histidine H493 to arginine in TDP1 greatly reduces the enzyme activity (Interthal et al. 2005).

Recent findings have shown that TDP2 is required for proper expression of TOP2β - dependent genes, many of which are implicated in normal neuronal development (Gomez-Herreros et al. 2014). Human individuals with impaired TDP2 activity show intellectual disability, seizures and ataxia. It was also demonstrated that TDP2 activity is required for normal levels of transcription of a set of androgen-receptor responsive genes in prostate cancer cells, even in the absence of exogenous TOP2 poisons (Gomez-Herreros et al.

2014). This finding also supports the idea of targeting TDP2 in cancer chemotherapy as hormone induced transcription of a set of genes heavily relies on TDP2 activity.

1.1.3.4 TDP2 as VPg unlinkase

It was recently demonstrated that TDP2 enzymatic function is hijacked by certain viral pathogens upon infection (Virgen-Slane et al. 2012). During their life cycle, the positive-strand RNA viruses of the picornavirus family (e.g. poliovirus and human rhinovirus) use a small viral protein (VPg) as a primer for RNA synthesis. This VPg needs to be removed upon polysome association. The enzyme which facilitates the 5'-tyrosyl RNA scission have been called VPg unlinkase but the true identity of this enzyme had been remained elusive for decades. By employing multi step purification, alongside with VPg unlinkase activity assays from uninfected HeLa cells, TDP2 was identified as VPg unlinkase. The finding that TDP2 is able to process protein-RNA complexes was further substantiated by biochemical assays and structural studies (Gao et al. 2014). The discovery of the involvement of TDP2 in host-pathogen interactions have shown another aspect of TDP2 function, and also identified TDP2 as potential drug target in viral infections (Caldecott 2012).

1.1.3.5 Structure-function examination of TDP2

Recent reports (Schellenberg et al. 2012) (Shi et al. 2012) have solved the structure and catalytic mechanism, of TDP2. A schematic overview of the domain organization of the molecule can be observed in Figure 1.1.5.

Shi et al. reported the structure of the full-length TDP2 from *C. elegans* (Shi et al. 2012). The structure is modular and composed of three parts: a forty amino acid long N-terminal stretch with no secondary elements followed by a small α-helical bundle composed of four helices (UBA-like domain), and the C-terminal catalytic domain. In the crystal structure the N-terminal part of TDP2 interacts *in trans* with the catalytic domain, resulting in two molecules in the asymmetric unit. The N-terminal stretch rich in glutamate and aspartate residues inserts into the DNA binding groove and the active site of the catalytic domain of an other molecule. This single stranded DNA mimicry suggests a specific type of regulation. The α-helical bundle makes contact with the catalytic domain via a hydrophobic interface. The multimerization of the full-length protein was also confirmed by small-angle X-ray scattering (SAXS) in physiological conditions. The proposed autoregulation of enzymatic activity by the N-terminal part of TDP2 remains to be investigated.

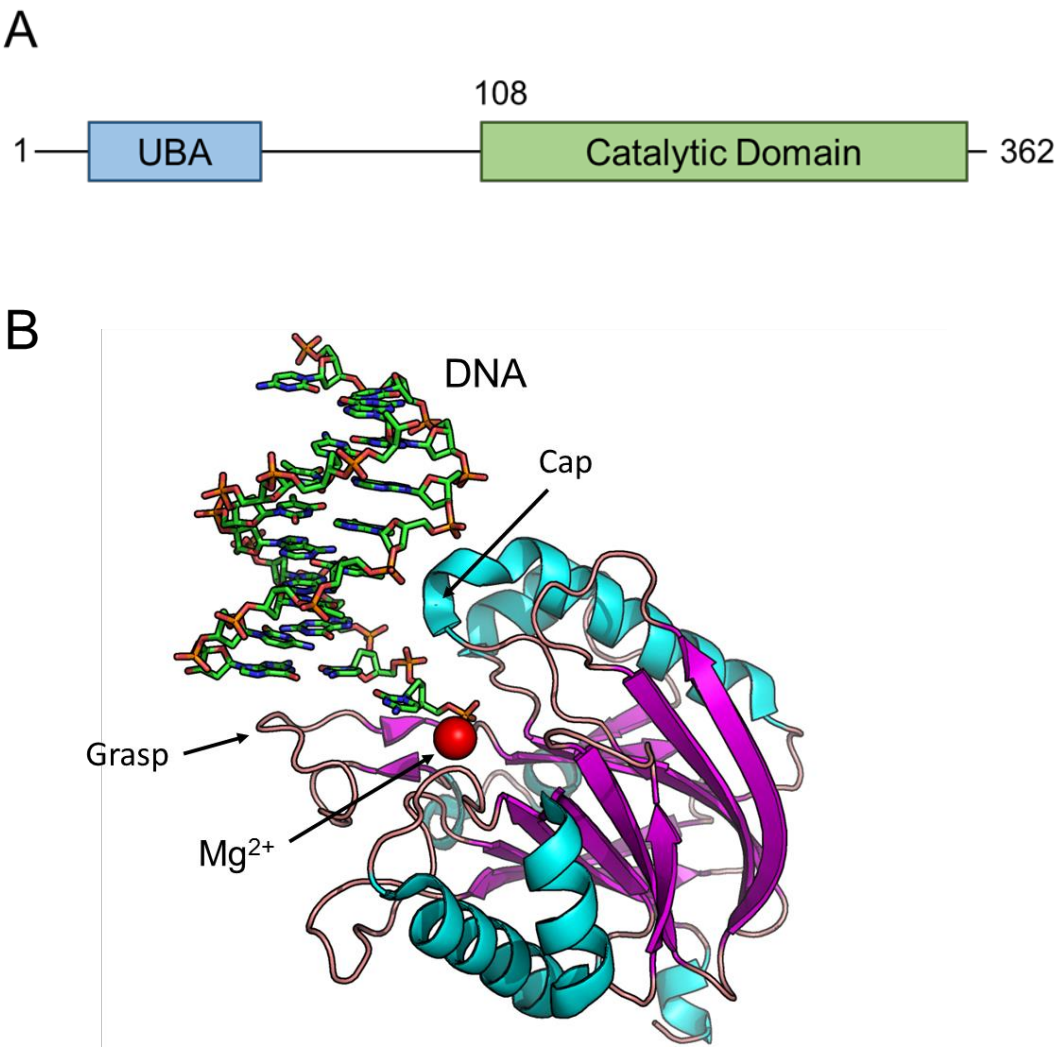

Figure 1.1.5. Domain structure of TDP2. **A.** Schematic representation of the domain organisation of TDP2. The N-terminal UBA domain is involved in the ubiquitination pathway and the C-terminal catalytic domain facilitates the scission of the phosphotyrosyl bond. **B.** Tertiary structure of the catalytic domain in complex with the cleavage product. The 5' end of the DNA is accomodated and oriented into the active site by the motifs 'Cap' and 'Grasp'. The magnesium cation (red sphere) is involved in phosphodesterase reaction. (Schellenberg et al. 2012)

An earlier report (Li et al. 2011) predicted that the N-terminal part of TDP2 encodes a ubiquitin-associated (UBA)-like domain. The classical fold of the (UBA)-like domain is composed of three helices and it is a common structural motif in enzymes involved in the ubiquitination pathway. It is also worth noting that the catalytic domain of TDP2 harbours a SUMO-interaction motif. Sumoylated and ubiquitinated TOP2 is targeted for proteolysis and it is hypothesised that these modifications also facilitate the recruitment of TDP2 to degraded abortive TOP2 cleavage complexes to liberate DNA ends ready for ligation (Gao et al. 2014).

The catalytic domain is responsible for the phosphodiesterase activity (Figure 1.1.5). Databank searching and sequence alignment identified six well-conserved motifs, which consigns TDP2 to the Mg^{2+}/Mn^{2+} dependent phosphodiesterase superfamily (MDP) (Rodrigues-Lima et al. 2001). Within the MDP superfamily, TDP2 is most closely related to the apurinic/apyrimidinic endonuclease APE1, which is involved in BER (Sukhanova et al. 2005).

According to structural studies by Schellenberg et al., the catalytic domain of the murine TDP2 has a globular α-β fold composed of a 12-stranded β-sandwich surrounded by eight α-helices (Schellenberg et al. 2012). The DNA is directed to the active site by the 'cap' and 'grasp' architectural motifs forming a cleft. The 'grasp' is composed of amino acids located on the M7-loop and mainly facilitates DNA binding by hydrophobic interaction. The 'cap' is located at the opposite site of the cleft; conserved amino acid residues making contact with the phosphate-sugar backbone of the DNA directing the 5' terminus into the active site. Based on structural and mutagenesis data, the mechanism by which TDP2 cleaves 5'-phophotyrosyl bonds is via single-metal-ion catalysis; Asp272 activates a water molecule for a nucleophilic attack on the 5'-phosphate. Mg^{2+} and three conserved residues (His236, His359 and Ser239) interact with the substrate phosphate to stabilize the pentacovalent transition state. The formation of a transient covalent complex was predicted in earlier reports as vanadate inhibits the catalytic activity of TDP2 reference (Zeng et al. 2011). It was also demonstrated that optimal Mg^{2+} concentration is essential for the catalytic activity of TDP2 (Adhikari et al. 2012). Elevated Mg^{2+} concentration has an inhibitory effect on TDP2 enzyme activity.

Interestingly TDP2 does not show structural similarities to the 3'-tyrosyl DNA phosphodiesterase (TDP1). TDP1 is a member of the diverse phospholipase D superfamily (PLD) (Interthal et al. 2001). TDP1 is a monomeric protein with two domains and contains two HKD (His, Lys and Asp) motifs which are responsible for the catalytic activity. The HKD motif is a common feature within the DLP superfamily and the catalytic reaction involves covalent complex formation between TDP1 and the 3'-end of DNA via a His residue of one of the HKD motifs. Structural studies of TDP1 complexed with vanadate (a common inhibitor of TDP1) have shown that vanadate is covalently linked the catalytic His residue (Davies et al. 2003).

1.1.3.6 TDP2 as potential drug target

The identification of TDP2 as a potential drug target is immediately suggested from its ability to resolve TOP2-DNA abortive complexes (Cortes Ledesma et al. 2009). TDP2 inhibitors might have utility in combination therapy with the TOP2 'poison' etoposide, which kills tumour cells by inducing TOP2 cleavage complexes. In addition, TDP2 inhibitors could be employed as a monotherapy in hormone-driven cancer types as their transcription program heavily relies on TDP2 activity (Gomez-Herreros et al. 2014). Unexpectedly, TDP2 was also identified as VPg unlinkase, an important factor in picornavirus infection (Virgen-Slane et al. 2012). TDP2 inhibitors might thus prove beneficial in the treatment of diseases such as meningitis, encephalitis, hepatitis or the common cold.

To identify small molecule inhibitors of TDP2, a high-throughput screening (HTS) strategy was developed (Thomson et al. 2013). Thomson and co-workers, in a collaboration between my laboratory and the Paterson Institute of Cancer Research, developed a chromogenic assay format, using 4-nitrophenyl phenylphosphonate (NPPP) as a chromogenic TDP2 substrate (Thomson et al. 2013). Approximately 100 000 compounds were screened in this HTS format, and toxoflavins and a singleton deazaflavin were identified confirmed hits (Raoof et al. 2013). Although toxoflavins showed clear structure activity relationship (SAR) for TDP2 inhibition, their redox liability hindered further exploration of this series. The single hit from a deazaflavin was developed into a promising series of compound showing clear SAR through a focused screen. These compounds had no observable redox liability, had good inhibition values and showed dosed dependent inhibition of TDP2 (Raoof et al. 2013).

1.1.4 The aim of Project I

My aim was to obtain structural information about TDP2 in complex with small molecule inhibitors to aid future drug development. In chapter 3 I will describe the early expression, purification and crystallisation trials of TDP2. In Chapter 4, the results of TDP2 binding and activity assays in the presence of small molecule inhibitors will be discussed and how these assays led to crystallisation trials which concluded in obtaining the first low-resolution dataset of a complex between TDP2 and an inhibitor. Chapter 5 describes the development of a surrogate protein in order to obtain high-resolution structure of these complexes, and Chapter 6 describes the results of initial cell-based assays to test the efficacy of these inhibitors.

1.2 Project 2 – Molecular analysis of the role of the XRCC1 BRCT1 domain

1.2.1 Single Strand Break Repair (SSBR)

DNA Single-strand breaks (SSBs) are the most common types of DNA damage and can arise at a frequency of tens of thousands per cell per day. These SSBs can arise directly by endogenous reactive oxygen species (ROS) and indirectly as intermediates of the base excision repair (BER) pathway (Pogozelski and Tullius 1998, Demple and DeMott 2002). SSBs can also arise from abortive Topoisomerase 1 (Top1) activity (El-Khamisy and Caldecott 2006). TOP1 releases torsional stress from DNA, which accumulates during transcription and DNA replication. If Top1 cleavage complexes are trapped by the presence of nearby lesions and/or collide with DNA or RNA polymerases, it can result in Top1-linked SSBs and DSBs, which if unrepaired it can lead to genetic instability or cell death. The most common pathway for repair of SSBs is single strand break repair (SSBR), which utilises four main steps: SSB detection, DNA end processing, DNA gap filling and finally DNA ligation (Figure 1.2.1) (Caldecott 2008).

In brief, poly(ADP-ribose) polymerase 1 (PARP1) detects SSBs and triggers the subsequent steps in the SSBR, by modifying itself and other proteins with long branched chains of poly(ADP-ribose) (PAR) (Kim et al. 2005). Autoribosylated PARP1 can promote the re-localisation of other repair proteins to damage sites (Pleschke et al. 2000). An important binding partner of PARP1 is the scaffold protein X-ray repair cross-complementing protein 1 (XRCC1) (El-Khamisy et al. 2003). XRCC1 orchestrates the subsequent repair steps by interacting and stimulating the different enzymatic components of SSBR and will be discussed in more detail in section 1.2.6. PARP1 is also implicated in DNA base excision repair pathway (BER), which is initiated by a DNA glycosylase that removes damaged or inappropriate bases by hydrolysis of the N-glycosidic bond (Durkacz et al. 1980, Krokan et al. 1997). The resulting apurinic-apyrimidic sites (AP sites) are cleaved either by AP endonuclease 1 (APE1), hydrolytically, or via the β-lyase function of a bifunctional DNA glycosylase (Wilson et al. 1997, Rykhlevskaya and Kuznetsova 2000).

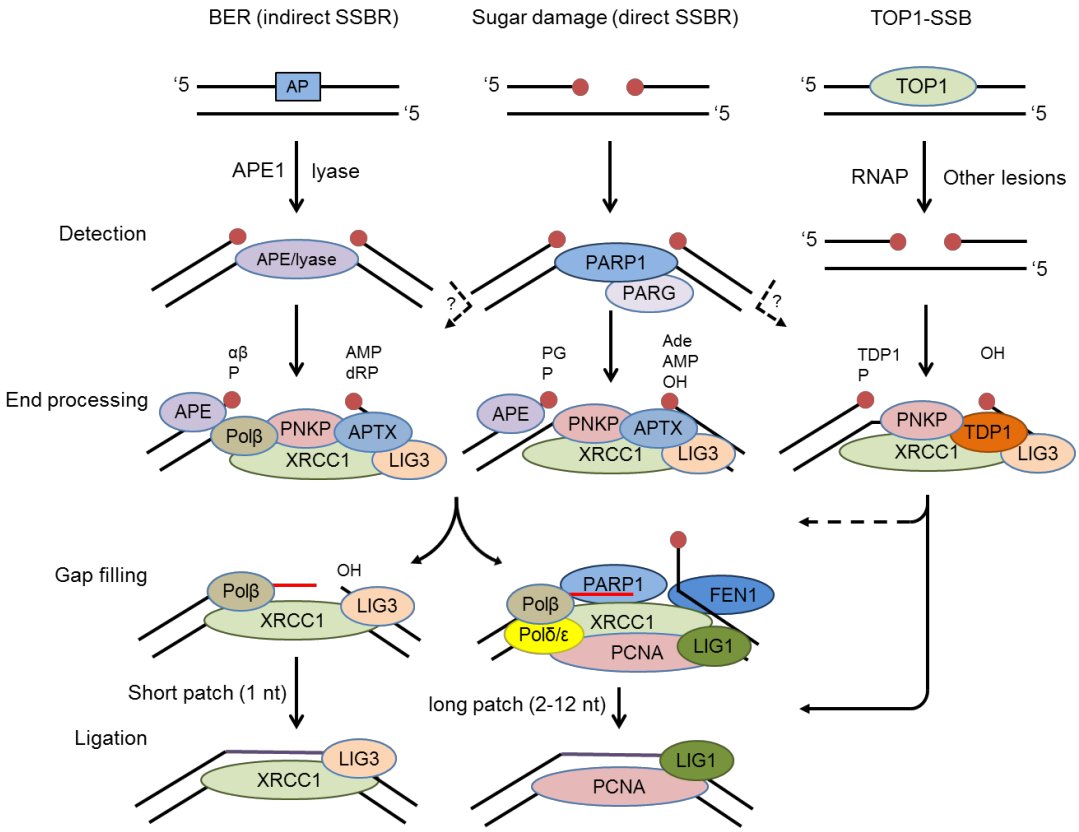

Figure 1.2.1. Schematic representation of single strand break repair. Single strand break can emerge directly or indirectly from apurinic/apyrimidinic (AP) sites removed by AP endonuclease I (APE1) or the lyase activity of bifunctional DNA glycosylases. TOP1-DNA adducts are mainly removed by Tyrosyl-DNA phosphodiesterase 1 (TDP1). The gap is detected by poly(ADP-ribose) polymerase 1 (PARP1). Activated PARP1 auto-ribosylates itself and other protein molecules, mainly histones. The ribosylation recruits downstream repair factors such as the scaffold protein XRCC1. When required, end processing enzymes (e.g. PNKP and APTX) restore the damaged ends to their canonical chemistry. During gap filling Polβ can insert a single nucleotide (short patch), or can synthesise a 10-12 nucleotide long flap which is then resected by Flap endonuclease 1 (FEN1) (long patch repair). The DNA ends are then ligated by either Lig3 (short patch repair) or Lig1 (long patch repair). KEY: RNAP, RNA polymerase, α, β, 3'-α, β unsaturated aldehyde; Ade, aldehyde; AMP, 5'-AMP; dRP, 5'-deoxyribose phosphate; OH, 5'-hydroxyl; P, 3'-phosphate; PG, 3'-phosphoglycolate. Figure is adapted from Caldecott, 2008 (Caldecott, 2008).

During DNA end processing the damaged 3'-and 5'-termini are restored to 3'-hydroxyl (3'-OH) and 5'-phosphate (5'-P) termini. The chemistry of these damage termini can be quite diverse, (e.g. 3'-phosphate, 5'-hydroxyl, 5'-aldehyde or phosphoglycolate) and so this step requires the availability of a broad range of enzymes (Figure 1.2.1). These enzymes were discussed in section 1.1.1.1. Once DNA termini are restored to their canonical chemistry, DNA gap filling can occur. This can involve either short patch or long patch gap filling (Frosina et al. 1996). Short patch gap filling involves the incorporation of a single nucleotide by DNA polymerase β (Polβ), whereas long patch gap filling involves the incorporation of two or more nucleotides, using Polβ, an/or Polδ/Polε (Polδ/ε). Flap endonuclease 1 (FEN1) is also required during long patch gap filling, to remove the displaced nucleotides as a 5'-single-strand flap (Liu et al. 2005). This reaction is stimulated by PARP1 and the proliferating cell nuclear antigen (PCNA) (Gary et al. 1999).

The final step of SSBR is DNA ligation by DNA ligases (Cotner-Gohara et al. 2008). DNA ligase III (Lig3) restores an intact phosphodiester backbone, thereby ligating the break, during short-patch repair step, whereas during long-patch repair SSBs are thought to be ligated by DNA ligase I (LIG1). However, recent reports suggest that there is much enzymatic redundancy in this step, with Lig1 and Lig3 fulfilling overlapping functions (Simsek et al. 2011).

1.2.2 ADP ribosylation is a post-translational modification

ADP-ribosylation is a common protein post-translational modification, and is triggered during DNA repair, transcription, and cell cycle progression (Schreiber et al. 2006, Leung 2014). Poly(ADP-ribose) polymerases (PARPs) synthesise PAR polymer at the expense of NAD^+ (Chambon et al. 1963). The cellular concentration of NAD^+ is 400-500μM in undamaged mammalian cells (Elliott and Rechsteiner 1975). However, during extensive DNA damage, the intracellular concentration of NAD^+ can decrease to 10-20% of the normal level (Goodwin et al. 1978).

1.2.2.1 Mono(ADP-ribosylation)

ADP ribosylation is phylogenetically well-conserved modification of target molecules (Ame et al. 2004). A subtype of this reaction is mono (ADP-ribosylation) and it is common in both eukaryotes and prokaryotes (Okazaki and Moss 1999). In prokaryotes, the best-studied mono-ribosylating enzymes are bacterial toxins, such as cholera toxin, heat-labile

enterotoxin and diphtheria toxin (Corda and Di Girolamo 2003). Cholera toxin targets the α-subunits of GTP-binding proteins resulting in constant activation and high level of cyclic-AMP, which stimulates the release of fluids and ions from intestinal epithelial cells. The main amino acids modified by enterotoxins are glutamate, aspartate, cysteine and asparagine. ADP ribosylation of the diphthamine residue (modified histidine residue) on the eukaryotic elongation factor 2 (eEF-2) by the diphtheria toxin results in translational inhibition (Mateyak and Kinzy 2013).

The sirtuin (SIRT) family is an intracellular mono (ADP-ribosylating) enzyme family and well conserved in eukaryotes, prokaryotes and archaea (Corda and Di Girolamo 2003). SIRT proteins regulate development, metabolism, heterochromatin formation, DNA transcription and DNA repair (Yamamoto et al. 2007). Most mammalian SIRTS are involved in histone deacetylation; however, SIRT2 is a known tubulin deacetylase. The family of the eukaryotic mono-ADP-ribosyltransferases (e-MARTs) possess extracellular mono(ADP-ribosylase) activity. e-MARTs are either cell membrane proteins or secretory proteins, targeting arginine or cysteine residues on extracellular proteins (Okazaki and Moss 1996).

There are members of the PARP [poly(ADP-ribose) polymerase] family (discussed in detail in section 1.2.3), which also possess mono(ADP-ribosylase) activity. PARP3 is a nuclear protein and involved in DNA repair and is stimulated by both DSBs and SSBs, but exhibits a preference for breaks with 5'-phosphorylated termini (Rulten et al. 2011, Langelier et al. 2014); PARP enzymes PARP6-8, PARP10-12 and PARP14-16 are also mono(ADP-ribosylase) enzymes (Gibson and Kraus 2012). Nuclear substrates for mono-ribosylation are mainly histones. H1 histones are modified on glutamate residues E2, E15 and E114/E115/E117 and on Arg33; H2B mono-ribosylated on Glu2 (Burzio et al. 1979, Ogata et al. 1980). Other studies indicated arginine residue modification on H2A, H2B, H3 and H4 histones (Golderer and Grobner 1991).

1.2.2.3 Poly (ADP-ribosylation)

ADP ribosylation of target molecules can result in long, branched chains of poly(ADP-ribose) (PAR). PAR polymer is a homopolymer of ADPr units linked via glycosidic ribose-ribose 1'-2' bonds (Miwa et al. 1979) (Figure 1.2.2). The forming polymer is negatively charged and can alter the activity of the modified protein through steric or electrostatic changes. The chain length can reach 200-400 units and branching can occur very

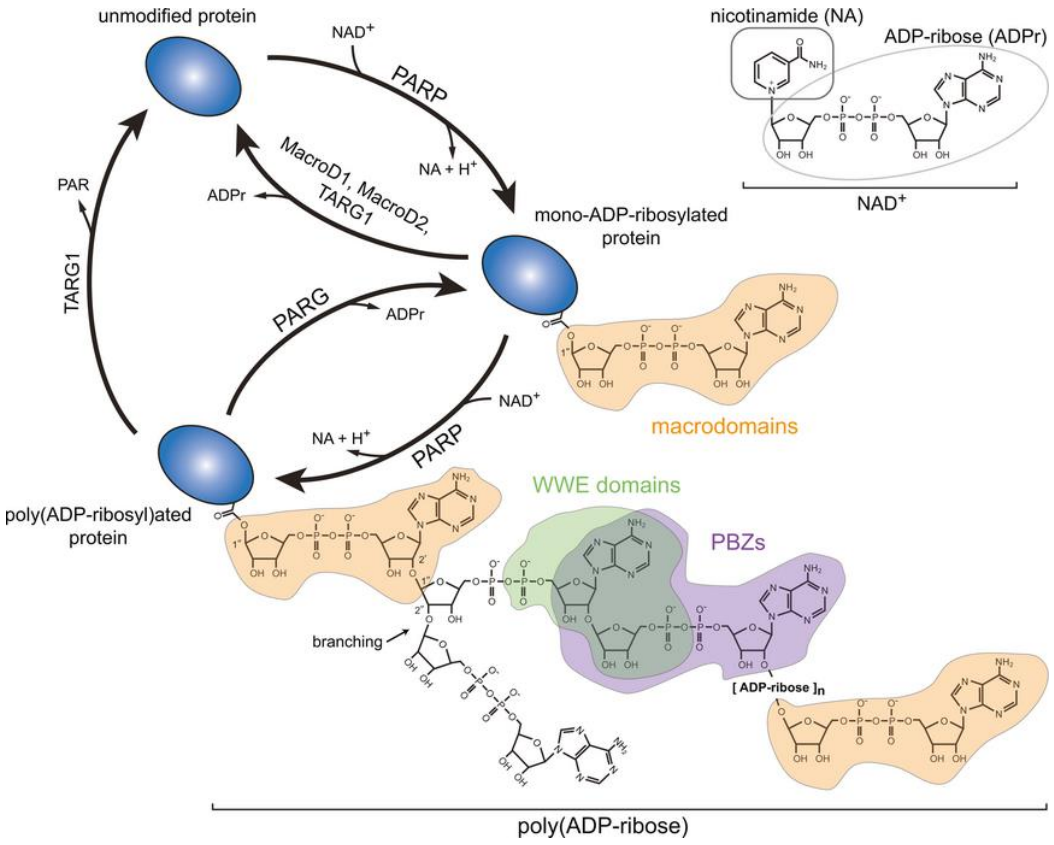

Figure 1.2.2. The cycle of ribosylation by PARP. PARP enzymes catalyse the transfer of an ADP-ribose onto a target molecule on the expense of substrate NAD^+. The PAR polymer can be recognised by several PAR binding motifs and domains. Macrodomains recognise terminal ADP-ribose, and the WWE domain recognises iso-ADP-ribose. PBZ domain facilitates PAR binding by stacking interaction with the adenine moiety and the glycosidic bond at the ADP-ribose-ADP-ribose junction. PAR chains are hydrolysed by PARG and ADP-ribosylhydrolyse 3 (ARH3), however, the most proximal ADP-ribose remains attached to the protein. Mono(ADP-ribose) is removed by MACROD1/2 and TARG1. Figure is adapted from (Barkauskaite et al. 2013).

irregularly (in every 20 or 50 elongation steps) (Kiehlbauch et al. 1993). During branching, the O-glycosidic bond forms between two nicotinamide-proximal ribose rings (N-ribose), whereas in the elongation step the incoming N-ribose binds to the adenine-proximal ribose of the growing polymer (Ruf et al. 1998, Otto et al. 2005). The structure of the polymer can be very complex and it was hypothesised that certain polymers can adopt helicoidal secondary structures (Minaga and Kun 1983). A proteome wide analysis of PARylated on aspartic and glutamic residues identified 340 target proteins (Zhang et al. 2013).

1.2.3 Poly(ADP-ribose) polymerase 1 (PARP1)

PARP1 is the founding member of the diverse poly(ADP-ribose) polymerase superfamily (PARPs), so far comprised of 17 proteins (Sousa et al. 2012). All PARPs share the conserved ADP-ribosyl transferase fold (D'Amours et al. 1999). Several PARPs have polymerase activity (PARP1, 2, 4 and 5a and 5b), the majority possesses mono-ribosylase function (see section 1.2.2.1) and two enzymes are enzymatically inactive (PARP9 and 13) (Gibson and Kraus 2012). PARP1 displays similar functional features as PARP2 and PARP3 (Ame et al. 1999, Boehler et al. 2011). The main roles of PARP1-3 in DNA damage responses are DNA damage sensing and signalling, DNA repair, and cell death induction. Some of these functions can overlap as *Parp1*$^{-/-}$ or *Parp2*$^{-/-}$ animals are viable, whereas *Parp1*$^{-/-}$/*Parp2*$^{-/-}$ mice exhibit embryonic lethality (Menissier de Murcia et al. 2003). PARP3 has a role in DSBR by recruiting APLF, which in turn increases the recruitment or the retention of XRCC4-LigIV at sites of DNA damage (Rulten et al. 2011). Apart from its well-established roles in DNA damage response, PARP1 also involved in transcription activation of genes by modulating the chromatin environment (Krishnakumar et al. 2008) and in the stress response (Luo and Kraus 2012).

PARP1 is 116kDa, and contains three major functional units: the N-terminal DNA binding region (NTR), the central BRCT domain (or auto modification domain, AMD) and the carboxy-terminal catalytic region (Figure 1.2.3) (Kreimeyer et al. 1984, Suzuki et al. 1987, Uchida et al. 1987, Ushiro et al. 1987). The DNA binding domain contains two tandem DNA binding zinc fingers, a third zinc binding domain that binds DNA and transmits allosterically binding to the catalytic domain (Langelier et al. 2008). PARP1 recognises and binds to single- and double-strand breaks and other aberrant DNA forms, such as cruciform DNA (Kim et al. 2004). Structural studies have shown that the first zinc finger and the second zinc finger from separate PARP1 molecules form a heterodimer aiding

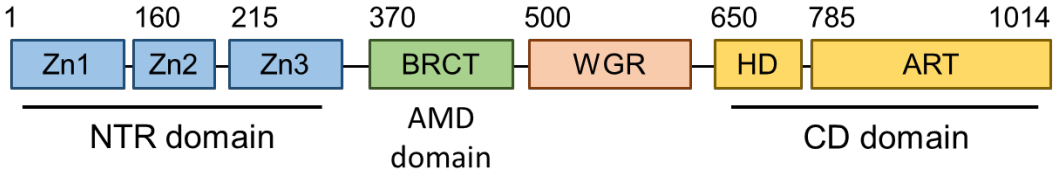

Figure 1.2.3. Schematic view of PARP1 domain organization. The N-terminal DNA binding domain (DBD) is composed of two zinc fingers (Zn1 and Zn2) and a zinc-binding domain. The central auto-modification domain (AMD) is a BRCT domain the main target for auto-ribosylation, and the well-conserved C-terminal catalytic domain (CD) is consisted of the WGR motif, the helical subdomain (HD) and the ART subdomain.

PARP1 dimerization and consequently trans-ribosylation (Ali et al. 2012). The central BRCT domain is the main PAR acceptor during auto- and trans-ribosylation, however, a recent study identified several ribosylation sites in the DNA binding region and in the catalytic domain as well. The main acceptors are aspartic and glutamic residues, although few lysine resides were also identified (Chapman et al. 2013) as PARylation sites.

The WGR (Trp, Gly and Arg rich) motif binds DNA and has an important role in mediating domain-domain contact essential in DNA dependent activity (Langelier et al. 2012). The conserved CD domain contains the helical subdomain (HD), and the NAD^+ binding ART subdomain [ART: (ADP-ribosyl) transferase]. ART contains the conserved glutamate residue (Glu 988) required for poly (ADP-ribose) polymerase activity (Marsischky et al. 1995). The E988Q mutation renders the protein to a mono-ribosylating enzyme.

Upon DNA binding the catalytic domain is activated and PARylates itself and a variety of other nuclear proteins such as DNA dependent protein kinase (DNA-PK), Topoisomerase I and histone molecules. The main target for PARylation is PARP1 itself. The rate and type of PARP1 auto-modification depends on the severity of the DNA damage. Mild or moderate stress leads to DNA response (Kim et al. 2005) and PARP1 functions to re-establish genome integrity. Severe PARP1 activation can lead to senescence or distinct cell death programs, such as necrosis, apoptosis or parthanatos (Ha and Snyder 1999, Gobeil et al. 2001). Hyper activated PARP1 can deplete cellular NAD^+ and ATP, which leads to necrosis. Activated PARP1 can also stabilise the tumour suppressor p53, which subsequently commence apoptosis (Soldani and Scovassi 2002) (Boulares et al. 1999). Activated caspases cleave PARP1 into inactive 24 and 81 kDa fragments, preventing the depletion of NAD^+ and ATP pools.

Parthanatos differs from apoptosis, necrosis and autophagy (Wang et al. 2009). When the accumulating DNA damage far exceeds the cell's repair capacity, PARP1 is over activated, which results in large scale PAR synthesis. If PAR production reaches a critical level it can lead to NAD^+ depletion. The overproduced PAR translocates from the nucleus to the cytosol and interacts with mitochondria. Apoptosis Inducing Factor (AIF) is released upon the interaction and translocates to the nucleus where it binds to DNA and triggers large-scale DNA fragmentation (Daugas et al. 2000).

Due to its role in DNA damage sensing PARP1 is an excellent chemotherapeutic drug target. Inhibiting PARP in homologous repair deficient cells (BRCA1 or BRCA2 deficient cells) leads to synthetic lethality by blocking base excision repair pathways (Helleday 2011). Olaparib and veliparib are potent PARP inhibitors and they are in advanced clinical trials. Studies indicate the besides their inhibitory function they trap PARP-DNA complexes which results in cytotoxicity (Murai et al. 2012).

1.2.4 PAR metabolism

The nature of the PAR polymers is transient due to their rapid degradation; the typical half-life of a polymer extends from 40 sec to 6 min (Boulikas 1992). The decomposition of PAR is dependent on length and branching; long, single-branched polymers degrade faster than short, multi-branched PARs (Li and Chen 2014). The main enzyme which degrades the PAR polymer by cleaving the glycosidic ribose-ribose 1'-2' bond is the poly (ADP-ribose) glycohydrolase enzyme (PARG) (Miwa and Sugimura 1971, Lin et al. 1997). PARG possesses both endo and exoglycolytic activity, and these activities can result in free, unbound polymer and ADPr, respectively (Braun et al. 1994). PARG is essential in the rapid turnover of PAR; PARG deficient mouse is embryonic lethal due to the uncontrolled accumulation of PAR. ADP-ribosylhydrolyse 3 (ARH3) is also a PAR processing enzyme, although less active than PARG (Mueller-Dieckmann et al. 2006, Niere et al. 2012). Both enzymes leave the most proximal ADPr attached to the target protein. The half-life of mono-ADP-ribosylation is significantly longer than PARylation. The single, most proximal ADPr moiety attached to the target molecule via an ester bond removed by the enzymatic activity of macrodomain containing enzymes (Jankevicius et al. 2013). Up to date there are 3 macrodomain enzymes described: Macrodomain-containing protein 1 and 2 (MACROD1 and MACROD2) and Terminal ADP–ribose protein glycohydrolase 1 (TARG1 or previously known as C6orf130). It was shown that homozygous deficiency in TARG1 can result in neurodegenerative disease (Sharifi et al. 2013).

1.2.5 PAR recognition motifs

Several different protein modules can facilitate PAR binding.

Poly(ADP-ribose) binding motifs (PBMs) are the most abundant PAR binding motif and are present in a broad range of proteins. The motif is approximately 20 amino acids long

and mainly composed of Lys and Arg clusters interspaced with hydrophobic residues. However, the mechanism/s of PAR recognition by these motifs are unclear. This motif has been identified in several DNA repair proteins, including s XRCC1, XPA, Ku70, and also on histones, p53, and apoptosis inducing factor (AIF) (Pleschke et al. 2000).

PAR binding zinc-finger (PBZ) domains are small (3.5kDa) modules, and first were found on the DNA damage response protein Checkpoint with Forkhead-associated and RING Domains (CHFR) (Oberoi et al. 2010). The fold is unordered and stabilized by a rigid zinc-coordinated backbone which facilitates PAR binding by stacking interaction with the adenine moiety and the glycosidic bond at the ADP-ribose-ADP-ribose junction. Tandem PBZ domains of APLF have affinities for ADP-ribose in the subnanomolar range (Eustermann et al. 2010) (Figure 1.2.4 A).

WWE domain was named after its most conserved residues. Iduna (RNF146) is a ubiquitin 3 ligase and its WWE domain facilitates PAR binding by recognising the iso-ADP-ribose moiety (Wang et al. 2012) (Figure 1.2.4 B). PARylated proteins can be targeted to proteosomal degradation by PAR dependent ubiquitylation (DaRosa et al. 2015). The specificity for iso-ADP-ribose over ADP-ribose shows preference for binding to PAR over mono-ADP-ribose. Other proteins harbouring the WWE domain are PARP enzymes (PARP11 and PARP14) and E3 ubiquitin ligases further supporting the idea that WWE domains link PARylation to ubiquitin pathways (He et al. 2012).

Macrodomain folds are found in all kingdom of life. The domain was named after macroH2A, a histone variant in which this domain fold was first identified. MacroH2A1.1 employs PAR binding to regulate chromatin remodelling (Timinszky et al. 2009). The domain fold is composed of a six-stranded β-sheet and five α-helices. The recognition of ADP-ribose is mainly facilitated by stacking interaction with the adenine ring, and additional interactions with the pyrophosphate and the distal ribose further increases the specificity (Figure 1.2.4 C). Macrodomains have been identified in several other proteins, and in particular in histone variants, PARP enzymes and chromatin remodelling enzymes. Some macrodomain containing proteins can also show catalytic activity and can remove the most proximal glutamate-linked ADPr unit, therefore fully reversing the PARylation of target proteins. MACROD1, MACROD2 and TARG1 (C6orf130) (Jankevicius et al. 2013) are all terminal glycohydrolases as it was discussed previously. PARG -the main PAR metabolizing enzyme- also harbours the macrodomain fold (Wang et al. 2014).

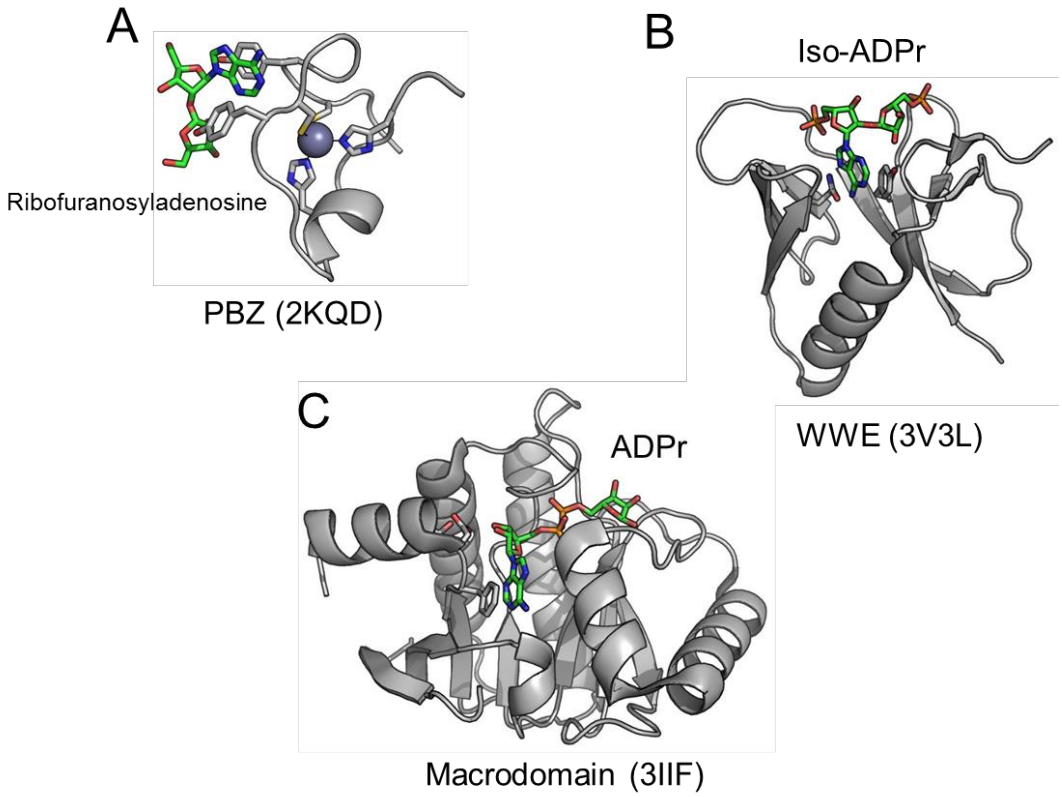

Figure 1.2.4 Molecular cartoon representation of known PAR binding modules. **A.** The first PBZ (Poly(ADP-ribose)-binding Zinc finger) domain of human APLF in complex with ribofuranosyladenosine; a PAR mimetic (PDB accession code: 2KQD) (Eustermann et al. 2010) **B.** WWE domain of RNF146 in complex with iso-ADPr (PDB accession code: 3V3L) (Wang et al. 2012) **C.** Macrodomain of macroH2A1.1 in complex with ADPr (PDB accession code: 3IIF) (Timinszky et al. 2009). ADPr and its mimetics are shown in stick representation (carbon atoms coloured green).

Forkhead-associated (FHA) domains were first identified in forkhead transcription factors (Hofmann and Bucher 1995). This highly conserved domain is a known phospho-protein binding module (Durocher et al. 1999) and has a strict preference towards phospho-threonine (pThr). However, it is able to recognise diverse patterns surrounding pThr (Mahajan et al. 2008). Proteins harbouring the FHA domain have been identified in eukaryotes and certain prokaryotes and these proteins are involved in many cellular processes, such as signal transduction, transcription, protein transport, DNA repair and protein degradation. The domain itself is composed of 80-100 amino acids and folds into an 11 stranded β sandwich, which can contain small helical insertions embedded between the loops connecting β strands (Durocher et al. 2000). The homology region of the FHA domain spans 8 β strands on one face of the β sandwich and harbours the phospho-peptide recognition motif. Phospho-peptide binding is mediated by the connecting loops of β3/β4 β4/β5 and β5/β6 (Durocher et al. 2000). FHA domain harbouring proteins also has important roles in DNA repair. The FHA domain of NBS1 and Ring Finger Domain 8 (RNF8) recognises the pThr motifs of MDC1, and these interactions facilitate the amplification of DNA damage signalling (Huen et al. 2007). Recently, the FHA domain of PNKP and APTX has been identified as PAR binding domain (Li et al. 2013). Li et al (2013) also demonstrated that similar to the phosphate-binding pocket identified on other FHA domains, two conserved lysine residues form the PAR-binding pocket in both PNKP and APTX. ITC assays have demonstrated that the FHA domain binds iso-ADPr and not ADPr. The FHA-BRCT fusion domain of NBS1 was also shown to interact with PAR, and mutational analysis has shown that the BRCT domain is responsible for the interaction (Li, Lu et al. 2013).

BRCT domains are denoted such from their identification at the C-terminus of the breast cancer type 1 susceptibility protein (BRCA1) (BRCT stands for BRCA1 Carboxy Terminal domain) (Koonin et al. 1996). Further screening in protein databases identified a great number of nonorthologous proteins containing this well conserved domain (Bork, Hofmann et al. 1997). BRCT domains are remarkably versatile and involved in many different interactions (Leung and Glover 2011). They can mediate protein-protein interactions by recognising phospho serine (pSer) motifs by their phosphate-binding pocket (Yu et al. 2003) formed of conserved lysine and serine residues. E.g. the MDC1 BRCT domain interacts with pS139 on H2Ax (Stucki et al. 2005). BRCT domains can appear in tandem in several proteins [i.e. the p53 binding protein (53BP1), Rad9 and

BRCA1] or in multiple copies separated by a loop (i.e. XRCC1) or as a single copy (DNA ligase 3 (LIG3) and PARP1) (Derbyshire 2002) (Nnakwe et al. 2009) (Taylor et al. 1998) (Loeffler et al. 2011). Tandem BRCT domains can achieve high selectivity towards phosphorylated proteins. E.g. in BRCA1 the N-terminal BRCT1 domain harbours the phosphate-binding pocket selective towards pSer, and a secondary recognition site is formed at the BRCT1/BRCT2 interface (Clapperton et al. 2004).

It was recently demonstrated that BRCT domains are also involved in poly(ADP-ribose) binding; e.g. BARD1 BRCT1 recognises PAR (Li and Yu 2013). A further study also identified the BRCT domain of Ligase4 as poly(ADP-ribose) binding module (Li et al. 2013). ITC assays have shown that the domain binds ADPr, and the interaction is mediated by conserved lysine and serine residues.

1.2.6 XRCC1

XRCC1 is a scaffold protein and its function is to facilitate SSBR by interacting with enzymatic components of the repair process. Cells lacking functional XRCC1 are hypersensitive to genotoxins (Thompson et al. 1982) (Shen et al. 1998). EM9 is a well-studied XRCC1-mutant Chinese hamster ovary cell line (CHO) in which there is increased sensitivity (10-fold) to alkylating agents such as ethyl methanesulfonate (EMS) or methyl methanesulfonate (MMS) (Cantoni et al. 1987). Reduced level of sensitivity caused by near-visible light, blue light and UV (Churchill et al. 1991). The rate of SSBR is reduced five-fold and the elevated level of unrepaired SSBs and if these unrepaired lesions collide with replication forks it can result in higher number of DSBs (Kuzminov 2001). XRCC1 is indispensable during embryonic development (Tebbs et al. 1999); and it is also required in neurogenesis of cerebellar interneurons (Lee et al. 2009).

XRCC1 is comprised of three domains connected by two linker regions (London 2015). The first interdomain region between the N-terminal domain and the BRCT1 domain harbours the nuclear localization signal (NLS) (Masson et al. 1998). This linker region has also been reported to be site of the interaction with Proliferating Cell Nuclear Antigen (PCNA) (Fan et al. 2004); PCNA is a homotrimeric loading clamp, implicated in DNA replication and also in DNA repair (Jaskulski et al. 1988, Shivji et al. 1992). The second linker region resides between the BRCT1 and BRCT2 domains and contains several casein kinase 2 (CK2) phosphorylation sites (Loizou et al. 2004) (Parsons et al. 2010). The

phosphorylation of XRCC1 by CK2 stimulates the binding and activity of PNKP, APTX and APLF via their FHA domain (Date et al. 2004, Iles et al. 2007, Ali et al. 2009).

Each domain interacts with different repair protein binding partners (Figure 1.2.5) (Caldecott 2003). The N-terminal domain (NTD) interacts with Polβ (Caldecott et al. 1996, Kubota et al. 1996, Gryk et al. 2002). The central BRCT1 domain is the most conserved region of XRCC1 and can be found in plants and insects. The structure of BRCT1 was solved by Nagashima et al, using NMR spectroscopy (PDB accession code: 2D8M). However, there is currently no published report available. The domain fold is a typical BRCT fold, a central β-sheet formed by 4 parallel β-strands flanked by 3 α helices. Due to the high number of basic amino acids, the domain has a p*I* of 10.2, which is consistent with its role in PAR mediated interactions. Pleschke et al identified a PAR-binding motif (PBM) at the C-terminal part of BRCT1, composed of mainly hydrophobic and basic residues (Figure 1.2.6 A) (Pleschke et al. 2000). However, a more recent report identified the phosphate-binding pocket as the site of PAR binding (Figure 1.2.6 B) (Li et al. 2013). The phosphate-binding pocket is a well-conserved motif in BRCT domains mediating phospho-serine binding, and mutating the conserved Lys369 in BRCT1 abolished PAR interaction. The main role of the BRCT1 domain is to mediate PAR dependent recruitment of XRCC1 to sites of DNA damage (Masson et al. 1998, El-Khamisy et al. 2003).

The structure of the C-terminal BRCT2 domain of XRCC1 was solved in 1998 (Zhang et al. 1998). BRCT1 and BRCT2 domains show great homology and 20% identity in amino acid composition. The BRCT2 domain forms a complex with Lig3α (Caldecott et al. 1995). As Lig3α completely lacks nuclear localization signal, its transport to the nucleus relies on the interaction with BRCT2 (Ellenberger and Tomkinson 2008). The BRCT2 domain of XRCC1 was shown to interact directly with DNA (Yamane et al. 2000), however, the physiological implication of this needs to be addressed. Other reports also suggested that XRCC1 preferentially binds DNA with single-strand breaks with a gap size of less than five nucleotides (Mani et al. 2004).

1.2.7 The aim of Project II

My goal was to investigate the interaction between the BRCT1 domain of XRCC1 and poly(ADP-ribose) applying cell biology and molecular biology techniques.

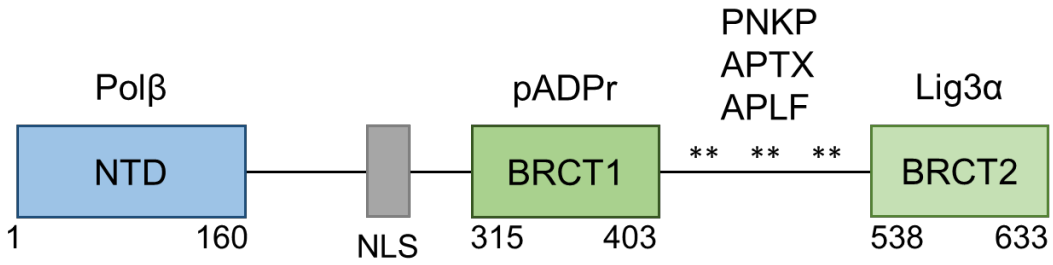

Figure 1.2.5. Schematic view of XRCC1 domain organisation. The N-terminal domain (NTD) interacts with Polβ. BRCT1 domain is the site for the PAR interaction and BRCT2 is the site of interaction with LIG3. The grey box represents the nuclear localization signal (NLS). The two interdomain linker regions between the three domains are also involved in protein-protein interactions. ** denotes phosphorylation sites by the protein kinase CK2.

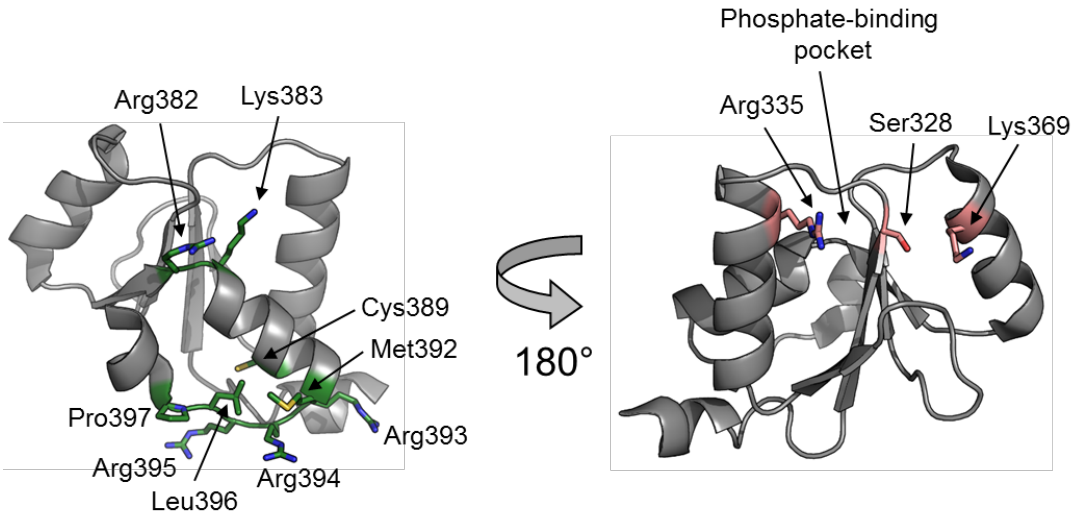

Figure 1.2.6: Ribbon diagram representing the structure of the BRCT1 domain of XRCC1. **A.** On the right: the hypothetical poly(ADP-ribose) binding motif composed of mainly basic and hydrophobic residues is depicted (in green). **B.** On the left: the phosphate-binding pocket is composed of well-conserved amino acid residues.

1.3 Introduction to X-ray Crystallography

1.3.1 Protein crystallisation

Through a controlled dehydration process, proteins can be forced to adopt a crystalline form. Crystallisation is thought to occur when the concentration of the protein reaches a metastable or 'supersaturated' state, which promotes the formation of micronuclei, from which a crystal is 'seeded' and can grow. This equilibrium process is described by the 'phase-diagram' depicted in Figure 1.3.1 (Chayen and Saridakis 2008).

More than 150 years after the crystallisation of haemoglobin was reported, the approach to crystallising proteins is still largely one of trial and error – as the successful formation of protein crystals is dependent on many different experimental factors, including: temperature, pH, salt concentration, the presence of other chemical agents, as well as the purity of the protein sample itself.

In 1958, John Kendrew and his team reported the structure of sperm whale myoglobin, the first protein structure to be determined, albeit at low resolution (6Å) (Kendrew et al. 1958). Two years later, Kendrew *et al*., reported a high resolution structure (2Å) of the same protein (Kendrew et al. 1960). No Protein Data Bank (PDB) entry represents the original myoglobin structure, since the databank was not established until 1971. However, in 1973, Watson and Kendrew deposited a refined model, at a resolution of 2.0 Å, available under accession code 1MBN.

1.3.2 Crystallisation methods

Several different methods have been developed, in order to overcome the inherent difficulties of crystallising proteins. Each method has its own distinct advantages, which may be useful for crystallisation certain types of proteins. A brief summary of the most common techniques is presented below.

1.3.2.1 Vapour Diffusion

Vapour diffusion is the most widely employed method of protein crystallisation. In this method a small droplet, created by mixing together protein and buffer/precipitant (crystallisation solution) is allowed to equilibrate against a reservoir containing a larger volume of the same mother liquor.

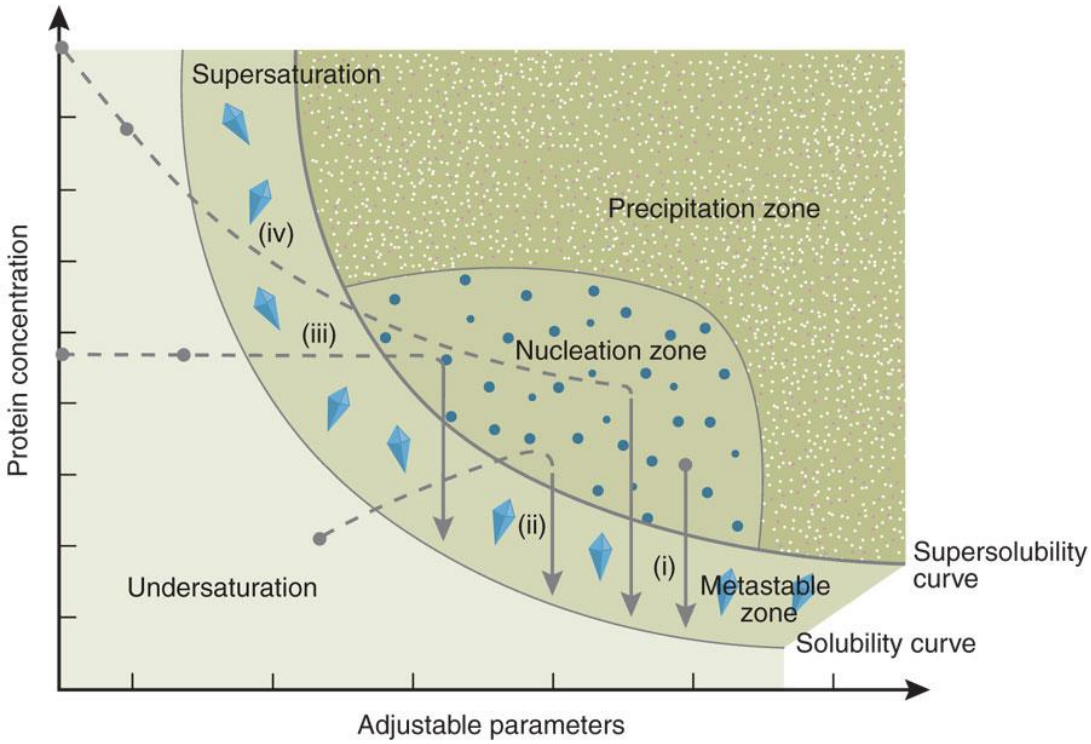

Figure 1.3.1. Phase diagram for protein crystallisation. Adjustable parameters (x-axis) include concentration of precipitation agent, pH or temperature. Solubility is defined as the concentration of protein in solution that is in equilibrium with crystals. The four major crystallisation methodologies are represented on the phase diagram: (i) Microbatch, (ii) Vapour diffusion, (iii) Micro-dialysis and (iv) Free-Interface Diffusion (FID). The starting condition for each method, and expected path through the phase diagram, are represented by the grey circle, and grey arrow respectively. Two alternative starting points are depicted for the dialysis and the FID methods, which depend on whether the protein was set up with precipitant or not. (Figure adapted from Chayen and Saridakis, 2008.)

Over time, water vapour is transferred from the droplet (less concentrated solution) to the reservoir (more concentrated solution), resulting in dehydration of the droplet, and an increase in both protein and precipitant concentrations. If the appropriate parameters are found for a given protein, crystal growth will occur in the droplet. The two most common vapour diffusion techniques are Hanging Drop and Sitting Drop (Figure 1.3.2).

1.3.2.2 Micro-batch

In the micro-batch method, the protein / crystallisation solution mixture is dispensed under low-density paraffin oil. The droplet is protected from evaporation, and from exposure to air (oxidation). In contrast to Hanging Drop, the concentrations of the protein or crystallisation solution will not markedly change over the time-scale of the experiment, due to the limited level of diffusion and evaporation through the oil layer.

1.3.2.3 Micro-dialysis

In the micro-dialysis method, the protein solution is separated from the crystallisation solution by a semi-permeable membrane, which allows the slow mixing and equilibration of chemical components, but prevents the diffusion of any protein molecules (Figure 1.3.2).

1.3.2.4 Free interface diffusion

In the Free interface diffusion method (FID), the protein solution and the crystallisation solution are in contact with each other, but are not pre-mixed. This is usually achieved in a sealed glass capillary. By equilibration and diffusion of molecules, across the protein/buffer interface, a concentration gradient is formed, which at the optimal level of supersaturation promotes nucleation and crystal growth.

1.3.3 Crystallisation screening methods

As the precise conditions to crystallise a given protein are still almost impossible to predict, several different screening methodologies have been developed. The most commonly employed methods are briefly described below.

1.3.3.1 Sparse Matrix Screening

The sparse matrix method uses ‘screens’ that are made up from collections of many different crystallisation solutions, proven to be successful in other crystallisation trials. Many sparse matrix crystallisation screens are commercially available from a wide range of suppliers.

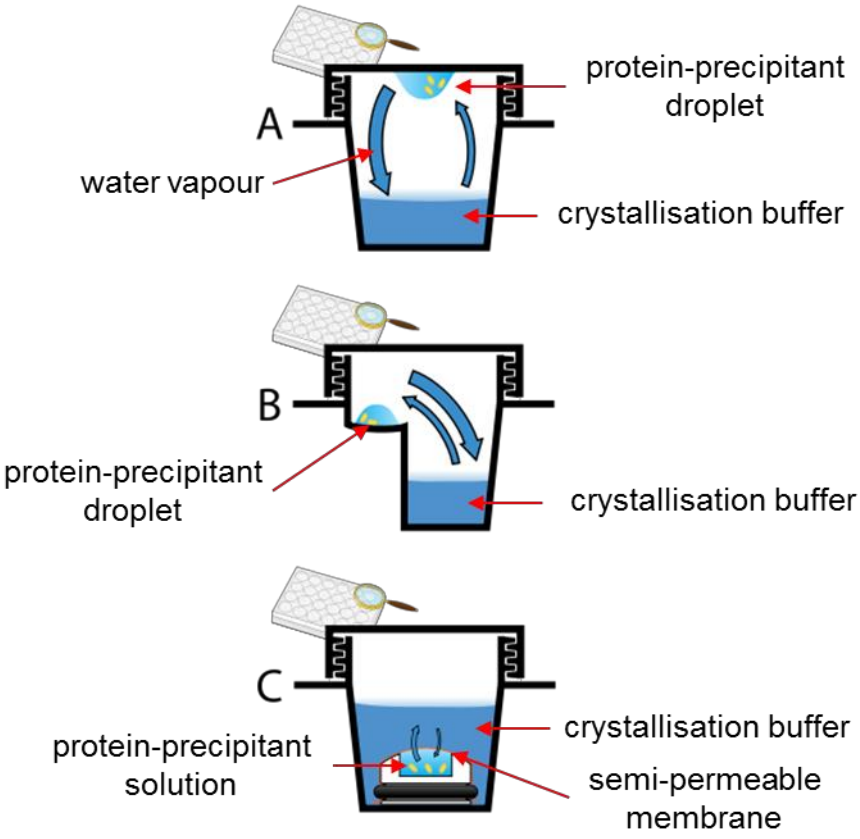

Figure 1.3.2. Diagram of different crystallisation set-ups. **A.** Hanging-drop **B.** Sitting-drop **C.** Microdialysis. Over time, water vapour is transferred from the droplet to the reservoir (large arrows) resulting in dehydration of the droplet, and an increase in both protein concentration and precipitant. If the appropriate parameters are found for a given protein, crystal growth will occur in the droplet. In **C.** the protein solution is separated from the mother liquor by a semipermeable membrane that allows slow equilibration of buffer components.

(Source: https://en.wikipedia.org/wiki/Protein_crystallization)

1.3.3.2 Grid screen

When a 'hit', i.e. crystals are formed, is found in sparse matrix screening, the crystallisation condition can be optimised (if required) by careful step-wise changes in the chemical parameters, e.g. one can screen the effect of pH and the concentration of a given buffer system, or salt concentration on crystallisation.

1.3.3.3 Additive Screens

Additive screens are typically libraries of small molecules and/or salts, which can affect protein solubility and crystallisation by manipulating protein-protein and protein-solvent interactions. Additive screens are also commercially available from many different suppliers.

1.3.4 Harvesting and Cryo-protection of Protein Crystals

Crystals are normally harvested from droplets with a 'cryo-loop', and then transferred into a suitable cryo-buffer, before either being frozen directly in a liquid nitrogen cryo-stream, or by plunging into in liquid N_2.

Cryo-protection of protein crystals greatly reduces the accumulation of radiation damage during data collection. During the cryo-protection process (as protein crystals contains substantial amounts of water) any ice crystal formation can physically damage the protein crystal.

A 'cryo-buffer' is usually a mixture of the crystallisation mother liquor and a cryo-protective agent, e.g. glycerol, glycol or sucrose, which prevents ice formation, and instead promotes a glass-like vitrification of the solution.

1.3.5 Crystals, space-groups and symmetry

Crystals are formed from individual molecules, which coalesce to form an ordered array, or repeating lattice. The repeated unit can contain multiple copies of the protein molecule and is termed the 'unit cell'. The unit cell can be described as a parallelepiped with lengths a, b and c, and angles α, β and γ. The internal symmetry of the unit cell can be described by its space-group.

There are 230 possible space-groups created using combinations of mathematical or 'symmetry' operations (rotation, translation, inversion and reflection). As proteins are intrinsically chiral molecules, the allowed symmetry operations are reduced to just rotation

and translation, leaving only 65 possible space-groups. Furthermore, there are only four possible rotational symmetries allowed in 3 dimensions: 2-fold (180º), 3 fold (120º), 4-fold (90º) and 6-fold (60º), as it is not possible to 'fill space' with the other symmetry types. This also results in 11 types of allowed 'screw-axes', where rotational symmetry operations are combined with translation.

The space-group is also defined by its crystal system and lattice type. A Bravais lattice (Auguste Bravais, 1850) is an infinite array of discrete points, generated by translation operations, and is described by the following equation:

$$\mathbf{R} = n_1\mathbf{a_1} + n_2\mathbf{a_2} + n_3\mathbf{a_3} \qquad \text{(equation 1.3.1)}$$

Where n_i is any value of integer and $\mathbf{a_n}$ is known as primitive vector.

In 3 dimensions, there are 14 Bravais lattice types: constructed by combination of the 7 crystal systems (triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal and cubic) and the five lattice centering operations (primitive, centered, face centered, internal and rhombohedral), which determine symmetry operators in the unit cell. The space-group is described by a letter denoting the lattice type, followed by the symmetry elements, represented by numbers with each axis with a symmetry element greater than 1 listed in order. For example: P 2_1 2_1 2_1, describes a primitive cell having a two-fold 2_1 screw axis on each unit cell axis.

The asymmetric unit (ASU) is the smallest unit of the crystal, which contains all the information required – when coupled with the symmetry operations defined by the space group – to mathematically reproduce the entire crystal lattice. Additional symmetry elements within the asymmetric unit can give rise to so-called 'non-crystallographic symmetry' but these symmetry elements are not defined by the space group of the unit cell.

1.3.6 X-ray diffraction

In this thesis, two different types of X-ray sources were used: (A) a MicroMax-007HF rotating anode source (Rigaku, Sevenoaks, UK), with a copper rotating anode providing monochromatic (single-wavelength) X-rays with a wavelength of 1.5418 Å. (B) Diamond Light Source synchrotron (DLS, Didcot, UK).

X-radiation (formed of X-rays) is defined as an electromagnetic radiation with wavelengths between 0.01-10 nm. This range of wavelengths matches the interatomic

distance of, and therefore can interact with, the electron cloud of atoms. Because of this, it is ideally suited for examination of molecules, where inter-atomic distances are on average around 0.15 nm (1.5Å) in length.

When an incident X-ray interacts with the electron cloud of an atom, its direction of propagation can be altered, though the process of elastic scattering. Elastic scattering conserves the kinetic energy of an incoming particle/wave, but modifies its direction of propagation. The scattered X-rays can also interact with each other, through either constructive or destructive interference.

During data collection the X-ray beam is directed through a collimator that renders the X-rays into parallel paths, producing a narrow beam. The diameter of beam can be further reduced by using focusing mirrors. The cryoloop holder containing the crystal is mounted onto a goniometer head, which help orient and move the crystal with great precision in order to stay in the path of the incoming X-ray beam during data collection. The crystal is kept at low temperature by nitrogen or helium stream to reduce radiation damage. The reflection intensities can be measured by detectors. There is a beam stop mounted beyond the crystal to prevent excessive radiation reaching the centre of detector. Detectors are scintillation counters counting incoming X-ray photons using phosphorescent material; the latest designs employ charge-couple devices (CCD camera).

In the crystal, each unit cell can be intersected by imaginary parallel planes defined by 'Miller' or lattice indices, termed *h*, *k* and *l*. The integers, represented by *h*, *k* and *l* denote where these planes intersect the unit cell axes a, b and c. Under certain conditions, a set of parallel planes in the crystal will interact to produce a 'Bragg peak' or 'reflection'. William Lawrence Bragg and William Henry Bragg demonstrated that such reflections are generated by constructive interference, but only occurs if the path difference between two parallel incident X-rays is equal to an integer multiple (n) of the wavelength (λ) of the incident radiation (Figure 1.3.3), where *d* is the inter-planar distance and θ is the scattering angle. Bragg's law or Bragg's diffraction is described by equation 1.3.2:

$$2d \sin\theta = n\lambda \qquad \text{(equation 1.3.2)}$$

Every atom of a given *hkl* crystal plane contributes to the reflection, and a reflection with an intensity of I_{hkl} is the result of the constructive interference of diffracted X-rays. The intensity of the reflection can also be described as a Fourier series. It is therefore possible

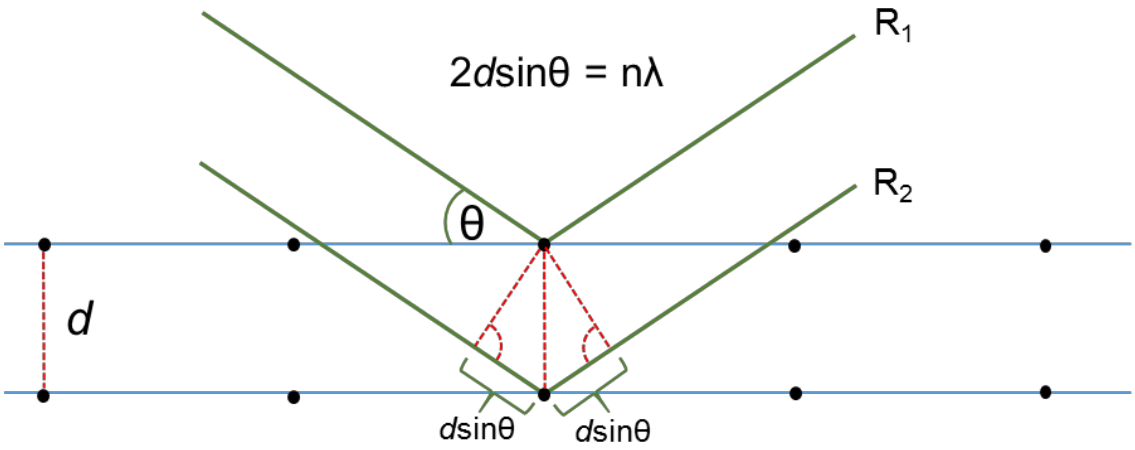

Figure 1.3.3. Diagrammatic representation of Bragg's law. Constructive interference only occurs when the pathway difference ($d\sin\theta$) of parallel incident X-rays (R_1 and R_2) is equal to an integer (n) multiple of wavelength (λ), and depends on the incident angle (θ) of the X-ray and the inter-planar distance (d). If the additional distance ($2d\sin\theta$) is not equal to the integral multiple of λ, destructive interference occurs.

to calculate the electron density at a given position (*xyz*) in the unit cell, by applying the following Fourier transform to *hkl* planes.

$$\rho_{(x,\,y,\,z)} = \frac{1}{V} \sum_{h}\sum_{k}\sum_{l} |F_{(h,k,l)}| \, exp\left[-2\pi i \, (hx + ky + lz - \alpha_{(h,k,l)})\right] \quad \text{(equation 1.3.3)}$$

Where: ρ(*xyz*) is the electron density at position (*xyz*), V is the unit cell volume, $|F_{hkl}|$ is the structure factor amplitude of the scattered X-ray and α_{hkl} is the associated phase.

The measured diffraction intensity is directly proportional to the square of the amplitude, i.e. $I_{hkl} = |F_{hkl}|^2$. As the intensities of each reflection can be measured during data collection, the amplitudes can be deduced, however, the phase information $\boldsymbol{\alpha_{hkl}}$ is not measured, and is lost. This is called the 'phase' problem.

There are several methods to recover the lost phase information. In this project, both single-wavelength anomalous dispersion and molecular replacement techniques were used in order to obtain phase information.

1.3.7 Obtaining Phases

1.3.7.1 Molecular replacement

Molecular replacement makes use of the 'Patterson function', in order to match a 'search' model, with known structure, to the 'target' molecules within the crystal lattice. Using the Patterson function does not require any *a priori* knowledge of phases (α_{hkl}). It is, however, calculated using the same Fourier summation (used to determine the electron density map) but only uses intensity information (I_{hkl}) rather than phases (equation 1.3.4).

$$P(u,\,v,\,w) = \frac{1}{V} \sum_{h}\sum_{k}\sum_{l} |F_{hkl}|^2 e^{-2\pi i (hu+kv+lw)} \quad \text{(equation 1.3.4)}$$

Where P(u, v, w) is the Patterson function, (u, v, w) coordinates locate a point in the Patterson map and $|F_{hkl}|$ is the structure factor amplitude.

The resulting 'Patterson map' is a set of position vectors describing the distances between each pair of atoms in the structure (relative positions). Vectors either correspond to pairs of atoms within the same molecule (self-Patterson vectors) are short, with their end points close to the origin of the Patterson map, or to pairs of atoms in symmetry-related molecules

(cross-Patterson vectors), which are longer, and with end-points far from the origin. Self-Patterson vectors calculated for both the search model and the target molecule are used to find the rotational relationship between the two sets of molecules, though a mathematical rotation function. The correctly oriented molecules are then matched by a mathematical translation function, using the cross-Patterson vector calculated between symmetry-related molecules. The correctly positioned model is then used to calculate initial phases, then further refinement using these phases and the experimental structure factor amplitudes $|F_{hkl}|$ can be used.

1.3.7.2 Anomalous scattering

When the energy (or frequency) of an incident X-ray photon is close to the 'absorption edge' of a particular atom, it is sufficient to promote an electron from the orbitals close to the atom nucleus (K or L-shell) directly to a higher shell, which has a higher average energy. These orbital alterations, result in the phenomenon of 'anomalous scattering' that causes small, but measurable differences in the intensity of reflections measured at planes h, k, l and -h, -k, -l (Friedel pairs) that are not normally present. Therefore, in crystals containing anomalous scatterers (atoms heavier than Sulphur) Friedel's law breaks down. Friedel's law states that in a centrosymmetric crystal the intensities of h, k, l and -h, -k, -l reflections are equal:

$$\text{Friedel's Law: } |F_{hkl}| = |F_{-h-k-l}| \text{ and } \alpha_{hkl} = \alpha_{-h-k-l} \qquad \text{(equation 1.3.4)}$$

Where: $|F_{hkl}|$ and $|F_{-h-k-l}|$ are the structure factor amplitudes, and α_{hkl} and α_{-h-k-l} are the phases of Friedel pairs.

The differences in Friedel Pair intensities, for anomalously scattering atoms, can be explained by an Argand diagram (Figure 1.3.4).

H, C, N and O atoms have negligible anomalous scattering in the range of X-ray wavelengths (0.35 – 3.5Å) typically used for diffraction experiments. However, good signals can be obtained at the K-edge of Selenium for example (Figure 1.3.5 A).

The difference in Friedel Pair intensities, can be used to locate the position of the anomalous scatterers in the unit cell, allowing phase information to be determined by either direct methods or by Patterson maps.

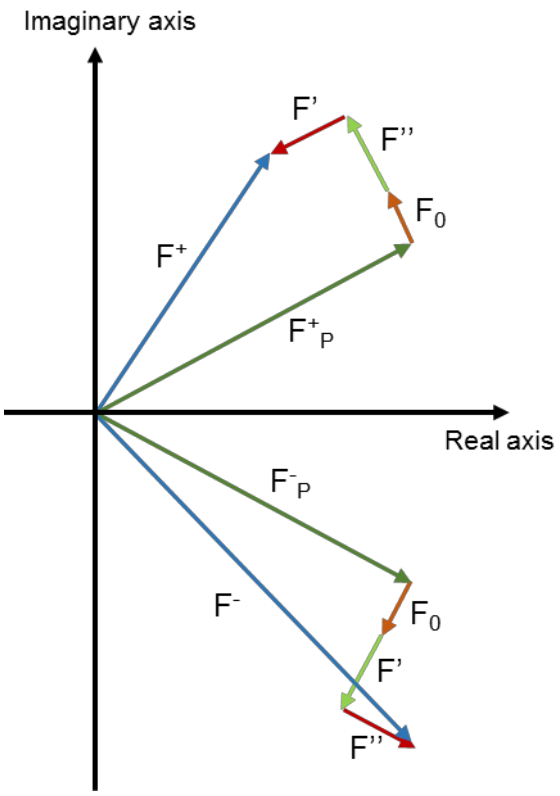

Figure 1.3.4. Argand diagram for Anomalous Diffraction. In the presence of an anomalous scatterer, Friedel's law breaks down: $|F_{hkl}| \neq |F_{-h-k-l}|$ and $\alpha_{hkl} \neq \alpha_{-h-k-l}$. Structure factors are represented by vectors, where the measured amplitude is represented by the vector length, and phase by the angle from the normal axis. The vectors coloured in dark green represent the structure factors emerging from non-anomalously scattering atoms; where the amplitudes are equal, but opposite in phase. The vectors coloured in dark blue represent the structure factors from anomalously diffracting atoms. The amplitudes are no longer equal, but still opposite in phase. The total structure factor, resulting from both sets of atoms is represented by summation of the displayed vectors: F_0 is the normal scattering term, F' is the dispersive term and F'' is the imaginary absorption term, shifted 90° in phase. This shift causes the differences in the amplitudes.

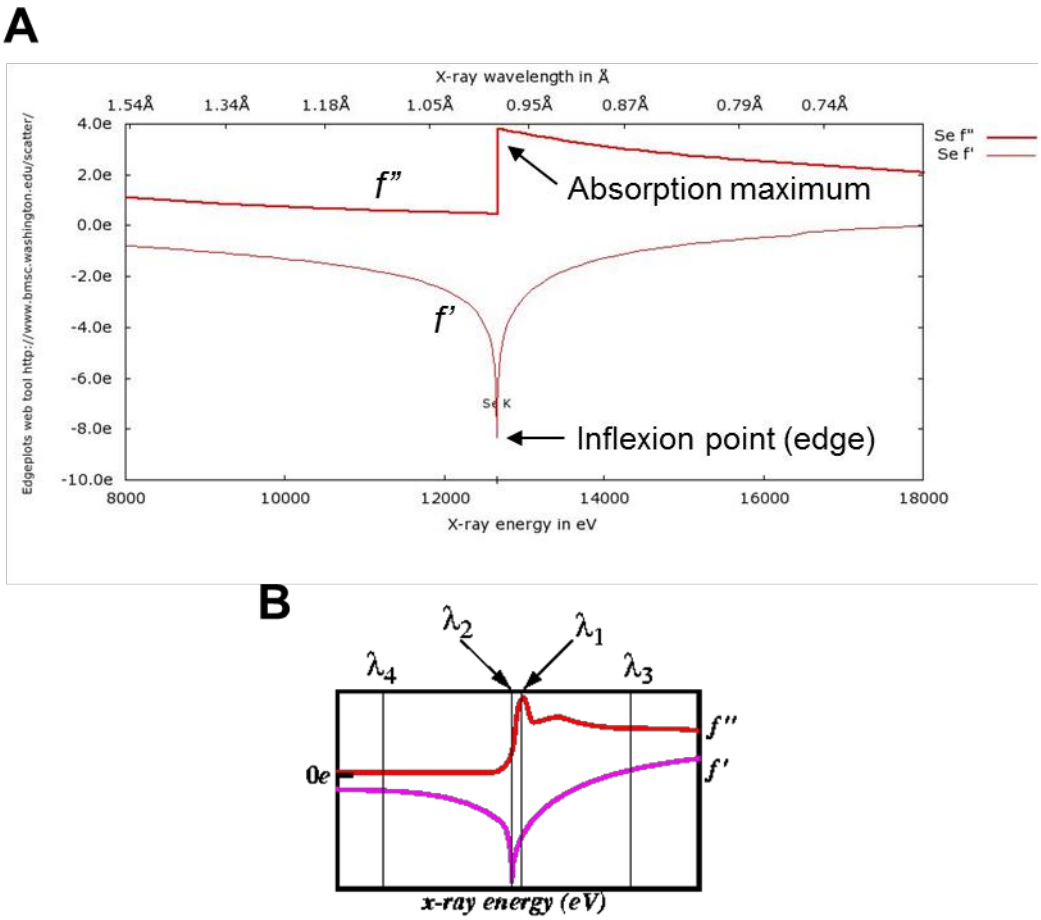

Figure 1.3.5. X-ray Absorption Spectra. **A.** Absorption spectra for the K-edge of selenium, showing scattering terms *f'* and *f"* at different energies (wavelength). **B.** Real (*f'*) and imaginary (*f"*) anomalous scattering factors vary greatly with X-ray wavelength. In Multiwavelength Anomalous Diffraction phasing, data sets are taken at the heavy atom absorption maximum (λ_1), at the inflexion point (λ_2), and at a remote wavelength (either at λ_3 or λ_4).

(Source: http://skuld.bmsc.washington.edu/scatter/AS_wavechoice.html)

1.3.7.3 Single-wavelength Anomalous Dispersion (SAD)

There are two main methods to measure anomalous diffraction. In Multi-wavelength Anomalous Diffraction (MAD) experiments, diffraction data are typically collected at three separate energies (wavelengths): a 'peak' dataset, calculated to give the largest anomalous signal (*f''*); an 'inflection' or 'edge' dataset, which is half-way down the *f''* absorption edge, i.e. at the inflection point of the curve, typically just a few eV less than the peak; and a 'remote' wavelength, either 'high' or 'low' depending on it is measured above or below the absorption edge (typically a difference of ~ +/-200 Ev) (Figure 1.3.5 B). Each measurement provides a distinct set of structure factors that can then be used to determine phase information.

In contrast, Single-wavelength Anomalous Dispersion (SAD) experiments employ just a single X-ray wavelength when collecting diffraction data. The incident wavelength is carefully chosen to produce the largest difference in Friedel Pair intensities (*f''*). The anomalously scattering atom substructure can be determined using a difference Patterson function using the amplitudes of the Friedel par. The total scattering of the protein can be calculated using equation (1.3.5):

$$F_{PA} = F_P + F_A + iF_A'' \qquad \text{(equation 1.3.5)}$$

Where F_{PA} is the total scattering of the protein and anomalously scattering atoms, F_P is the normal and dispersive elements from the scattering atoms, iF''_A is the imaginary phase component.

It is worth noting that two sets of phases, with opposite 'hands', are calculated from this type of experiment. This phase ambiguity is typically 'solved' by examination of the 'quality' of resulting electron density maps after density modification.

Using SAD, and modern experimental set-ups, it is now possible to measure the relatively weak anomalous scattering emerging from sulphur atoms (i.e. from cysteine and methionine side chains of a protein), and use this to phase x-ray diffraction data; the so-called 'native-SAD', or 'sulphur-SAD' method.

1.3.8 Data collecting and processing

Preliminary to data collection, the properties of a crystal are first analysed by collecting two to three small wedges of diffraction, typically separated by 30-90°, in order to allow 'indexing' – the process of determining the Bravais lattice (point group) of the crystal, its unit cell dimensions, as well as an estimate of the resolution of diffraction and mosaicity. Estimates of X-ray exposure time and optimal detector distance can also be obtained at this stage. With this information, an experimental strategy can then be determined, which allows a complete diffraction dataset to be collected with at least 2-fold multiplicity (2 measurements of each reflection). Proprietary software is generally available at synchrotron light sources, i.e. GDA (Diamond Light Source, Didcot, UK) or MXCube (European Synchrotron Radiation Facility, Grenoble, France) to facilitate this process.

Other computer software suites, such as those provided by the Crystallographic Computing Project 4 (CCP4, (Winn et al. 2011)) can also be used for data processing and manipulation, e.g. when diffraction data has been collected at a home source, or re-processing – for example, to alter point group assignments or resolution limits – is necessary/required.

In this thesis, either iMosflm (Leslie 2006), XDS (Kabsch 2010) or the xia2 automated pipeline (Winter 2010) were used for indexing and integration of all diffraction data.

Once indices have been assigned, the intensity of each reflection on each diffraction image can be measured. The collated data is then merged, scaled and truncated in order to produce a single experimental data file (MTZ file or equivalent), typically using the Pointless / Aimless / Ctruncate pipeline of the CCP4 suite (Winn et al. 2011).

Briefly, the merging processing involves the averaging of reflection intensities that are directly equivalent by crystal symmetry, as well as the summation of reflection intensities that have been recorded over several frames (so-called partials). Scaling then attempts to put all these observations onto a common scale, whilst also correcting for errors and inconsistencies caused either by the instrument or the crystal itself. The truncation process then converts the measured intensities into structure factor amplitudes ($|F|$).

The final output includes the averaged intensities (I) as well as structure factors (F), a significance score (estimate of uncertainness in the measured intensity), the completeness of the collected dataset, the mean signal to noise ratio mean ($I/\sigma I$) and the statistical

parameters, R_{merge} and/or R_{pim} which aim to show how unique reflection measurements agree/deviate from one another.

1.3.9 Model building, refinement and validation

After obtaining initial estimates of phases by either molecular replacement or by SAD/MAD experiments — the agreement between the measured and calculated amplitudes (from the initial model) can be improved by adjusting/moving the atomic coordinates of the model in small iterative steps. Restraints/constraints of these movements are based on theoretical and observed parameters, pertaining to known bond lengths, angles, stereo-chemical and conformational properties, which allow the model to be altered to better fit the electron density map, but to maintain expected/ideal geometries. The new model, and thus improved phases, can be then be used to generate a new electron density map – which can then itself facilitate further improvements to the model to be made – i.e. an iterative process, tending towards eventual convergence.

During model building and refinement, the agreement between the calculated model and the measured crystallographic data is validated by calculating an R-factor (R_{work}) (equation 1.3.6):

$$R_{work} = \frac{\Sigma hkl||Fobs|-|Fcalc||}{\Sigma hkl|Fobs|} \qquad \text{(equation 1.3.6)}$$

Where $|Fobs|$ is the observed structure factor and $|Fcalc|$ is the calculated dataset.

A low R-factor means that there is a good agreement between the model and the recorded dataset. A desirable R-factor for a protein model at 2.5Å resolution is ≤ 20%; which is in contrast to crystal structures of small organic molecules, at sub-1Å resolution, which can typically can be refined to R_{work} < 0.5%. The poor modelling of bulk solvent (water) in macromolecular crystallography is thought to be largely responsible for this discrepancy in R-factor values.

To avoid over-modelling, cross-validation is carried out throughout the refinement process, by calculation of R_{free} factor, where a small random subset of data is excluded from the refinement process (typically 5% of the total number of unique reflections). The value of R_{work} and R_{free} factors should be different by a factor of around 5%; any larger discrepancy indicates over interpretation of the electron density / model. During the refinement process

root-mean-square (rms) deviations of both the model's bond lengths and angles are also typically monitored.

In this thesis, either REFMAC5 (CCP4 software suite (Murshudov et al. 2011)) and/or phenix.refine (Adams et al. 2010) were used to refine the crystallographic model. The program COOT (Emsley et al. 2010) was used for model inspection and manual building into electron density maps.

REFMAC5 refinement typically consists of 10 cycles of maximum likelihood refinement to produce a refined model as well as two types of electron density map, the weighted map of $2m|Fobs|$ - $D|Fcalc|$ and $|Fobs|$ - $|Fcalc|$. The two map types guide model building in COOT; extra electron density in the difference map of $|Fobs|$ - $|Fcalc|$ can demonstrate the necessity for additional components in the model. The model then once again is refined in REFMAC. These steps were repeated until the model could not be improved further, and no improvement in R_{work}/R_{free} was observed; i.e. the refinement had reached convergence.

Models were also analysed with MolProbity, throughout the refinement process (Chen et al. 2010) to assess the quality of the model based on covalent-geometry and torsion-angle criteria. The MolProbity score represents the central protein quality statistics combining percentage Ramachandran not favored and bad side-chain rotamers. MolProbity also compares the model's geometry to other models with similar resolution in the PDB.

CHAPTER TWO

Materials and Methods

2.1 DNA cloning

2.1.1 Bacterial transformation

50µl chemically competent *E. coli* cells were thawed on ice, then incubated for 30 min in the presence of 1µg plasmid DNA. Cells were heat-shocked for 30 sec at 42°C, then transferred back to ice for an additional 2 min. 300µl antibiotic-free LB (Luria-Bertani) media was added to the cells, which were then incubated at 37°C, 220 rpm in a C25 incubator shaker (New Brunswick Scientific, Edison, NJ, USA) for an hour. Cells were then plated on LB-Agar plates (10g Tryptone, 5g Yeast extract, 10g NaCl and 15g Agar in 1L distilled water) supplemented with the required antibiotics for selection (e.g. 50µg/ml ampicillin) and then incubated overnight at 37°C in a static incubator (Heraeus Incubator, Thermo Fisher Scientific, East Grinstead UK).

2.1.2 Plasmid DNA preparation

To prepare plasmid DNA, a single transformed bacterial colony was used to inoculate 5ml LB media supplemented with the required antibiotics. The culture was incubated overnight at 37°C, 220 rpm in a C25 incubator shaker (New Brunswick Scientific). The next day, cells were harvested by centrifugation, at 4500 rpm for 10 min at room temperature, using an Allegra X-22R Centrifuge (Beckman Coulter, High Wycombe, UK), equipped with a SX4250 rotor. Pelleted cells were either stored at -20°C until required, or used directly for plasmid extraction. A QIAprep Spin Miniprep Kit (Qiagen, Manchester, UK) was used to isolate plasmid DNA, following the manufacturer's 'centrifuge' protocol. The purity and concentration of the plasmid DNA was measured using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). Purified plasmid DNA was stored at -20°C.

2.1.3 Polymerase Chain Reaction (PCR)

All PCR reactions were performed in a *MWG Biotech Primus* 25 thermal cycler. DNA fragments were amplified with KOD Hot Start DNA Polymerase (Merck Millipore, Feltham, UK) using dNTPs and buffers provided. All PCR primers were purchased from Eurogentec (Southampton, UK). Reaction mixes were typically composed of 10ng template DNA, 0.1µM forward and reverse primers, 200µM dNTP mix, 5µl 10x reaction buffer, 0.5µM $MgSO_4$ and 1µl KOD enzyme. Ultrapure water (supplied) was added to the reaction to a final volume of 50µl. PCR programs are provided in the relevant sections.

2.1.4 Agarose gel electrophoresis

Agarose (Genetic Analysis Range, Thermo Fisher Scientific) was dissolved in 1xTAE buffer in a microwave, and then supplemented with ethidium bromide (Sigma-Aldrich, Gillingham, UK) at a final concentration of 0.2µg/ml. The dissolved agarose was poured into a Horizon 58 gel casting deck to set (Thermo Fisher Scientific).

1x DNA loading buffer, from a 6x stock (0.15% w/v Bromophenol blue, 0.15% w/v Xylene cyanol and 30% v/v glycerol in water) was added to each DNA sample, before being loaded onto the agarose gel. The final concentration (w/v) of the agarose was determined by the expected size of the DNA sample. DNA samples were separated by electrophoresis at 120V for 30 min using a Horizon 58 Horizontal Gel Electrophoresis System (Thermo Fisher Scientific). A molecular mass standard (GeneRuler 1kb Plus DNA ladder, Thermo Fisher Scientific) was run alongside the samples in order to determine the size of the DNA analysed.

2.1.5 Agarose gel extraction

Following electrophoresis, DNA migrating at the desired molecular size was excised from the gel and either stored at -20°C or processed immediately. To obtain purified DNA, an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, UK) was used, following the manufacturer's recommended protocol.

2.1.6 TOPO cloning

2µl (~200ng) insert DNA was mixed with 1µl pCR4 Blunt TOPO vector (Thermo Fisher Scientific), 1µl supplied TOPO salt solution (1.2M NaCl, 60mM $MgCl_2$) and 2µl water, to give a total volume of 6µl. The reaction was incubated for 30 min at room temperature, before being used to transform a suitable strain of *E. coli*.

2.1.7 Restriction digest

Restriction enzymes were purchased from either Roche Life Science (Mannheim, Germany) or New England Biolabs (NEB, Hitchin, UK). For double digests, a restriction buffer compatible with both enzymes was chosen according to the manufacturer's recommendations. Typically, ~300ng of plasmid DNA was mixed with 0.5µl of each restriction enzyme, 2.5µl 10x restriction enzyme buffer, and water added to give a total volume of 25µl. The reaction was incubated for 60 min at 37°C.

2.1.8 DNA ligation

To prevent self-ligation, the 5' ends of the digested DNA vector were treated with Calf Intestinal Phosphatase (CIP, from NEB) prior to the ligation reaction. The phosphatase reaction was typically carried out during the restriction digest by adding 0.5µl CIP to the reaction mixture.

Gel-extracted insert and CIP-treated vector, at a molar ratio of 3:1, were mixed with 2µl 10x T4 DNA ligase buffer (NEB), 1µl T4 DNA ligase enzyme (NEB) and ultrapure water to a total volume of 20µl. The mixture was incubated at room temperature for 30 min. 5µl volume of the reaction was used to transform a suitable strain of *E. coli* competent cells.

2.2 Protein expression and purification

2.2.1 General Procedure

Transformation of *E.coli* strains, suitable for protein expression, follows the same protocol as previously described (Section 2.1.1). Once transformed, cells were streaked onto LB-agar plates supplemented with appropriate antibiotic(s) and then incubated overnight at 37°C, in a static incubator. All liquid cultures were grown in either Minitron or Multitron shaking incubators (Infors HT, Bottmingen, Switzerland).

Cells from 5ml and 50ml small-scale cultures were harvested by centrifugation in an Allegra X-22R Centrifuge (Beckman Coulter) using a SX4250 rotor, at 4500 rpm for 10 min at 4°C.

Cells from large-scale cultures (>1L) were harvested by centrifugation in an Avanti J-26XP centrifuge (Beckman Coulter) using a JLA8.1 rotor, at 5500 rpm for 10 min at 4°C. Harvested cells were either used immediately or stored at -20°C until required.

In general, cell pellets were thawed on ice, and re-suspended in 'Lysis buffer'. They were then disrupted by sonication, using a Sonics Vibra-Cell VCX 500 ultrasonic processor equipped with a large parallel probe (VWR, Lutterworth, UK). Cell debris and insoluble material were removed by high-speed centrifugation in an Avanti J-26XP centrifuge (Beckman Coulter) with a JA-25.50 rotor, at 20,000 rpm for 1 hour at 4°C. Where required, the resulting cell lysate was filtered through a 0.5µm Millex Syringe Filter unit (Merck Millipore).

For SDS-PAGE analysis, samples were applied to Novex 4-12% Tris-Glycine Protein Gels (Thermo Fisher Scientific) or TruPAGE precast gels (Sigma-Aldrich), and separated by electrophoresis for 35 min at 180V. Either Precision Plus Protein Standard (Dual Color, from Bio-Rad, Hemel Hempstead, UK) or SeeBlue Plus2 Protein Standard (Thermo Fisher Scientific) were used as molecular mass markers. An ÄKTApurifier 10 Protein Purification System operated by UNICORN 5.11 software (GE Healthcare) was used for all protein purification steps. Purified proteins were concentrated using Vivaspin 20 (5,000 and 10,000 MWCO) centrifugal concentrators (Sartorius Stedim), by centrifugation in an Allegra X-22R Centrifuge (Beckman Coulter), using a SX4250 rotor, at 4000 rpm and 4°C, until the desired volume was attained. The concentrated sample was then aliquoted, flash-frozen in liquid N_2 and then stored at -80°C until required.

2.2.2 Bacterial protein expression constructs

pTWO-E is an in-house modified form of pET-17b, encoding an N-terminal human rhinovirus-3C cleavable (HRV-3C) His_6 affinity tag. pTHREE-E is an in-house modified form of pGEX-6P-1, encoding an N-terminal HRV-3C cleavable, glutathione-S-transferase (GST) affinity tag. Both vectors were kindly provided by Dr Antony W. Oliver, University of Sussex.

2.2.3 Bacterial protein expression

Chemically competent *E.coli* were transformed according to section 2.1.1. For expression, strains BL21(DE3), BL21(DE3)pLysS (Promega, Chilworth, UK) and Rosetta2(DE3) pLysS (Merck Millipore) were used. For Selenomethionine labelling the auxotroph B834(DE3) (Novagen) *E. coli* strain was used.

2.2.4 Protein expression in insect cells

For protein expression in insect cells the Bac-to-Bac Baculovirus Expression System (Thermo Fisher Scientific) was used. Cultures of serum-free adapted Sf9 cells were grown in Sf-900II SFM media, supplemented with 5µg/ml Penicillin and Streptomycin, in 1L polycarbonate Erlenmeyer Flasks (Corning, USA) in a Multitron shaking incubator (Infors HT) set at 27 and 150 rpm. Infected cells were harvested by centrifugation, in an Avanti J-26XP centrifuge (Beckman Coulter) with a JLA8.1 rotor, at 1000 rpm for 10 min at 4°C. Cell counts and cell viability were determined by a Countess Automated Cell Counter

(Thermo Fisher Scientific) using a Trypan Blue staining method (Thermo Fisher Scientific).

DNA inserts were subcloned into the multiple cloning site (MCS) of pTWO-B at the *Nde*I and *Eco*RI restriction sites, using standard cloning procedures (see above). pTWO-B is a modified form of the pFastBac1 vector (Thermo Fisher Scientific) encoding an N-terminal HRV-3C cleavable, His_6 affinity tag. The vector was kindly provided by Dr Antony W. Oliver, University of Sussex.

The initial transformation of the DH10Bac strain of *E.coli*, and the generation of recombinant bacmid, followed the manufacturer's recommended protocol. Confirmed recombinant bacmids were resuspended in 40µl ultra-pure water and stored at -20°C until required.

Recombinant bacmid DNA was analysed by PCR using KOD Hot Start DNA polymerase (Merck Millipore). pUC/M13 forward (5′-CCCAGTCACGACGTTGTAAAACG-3′) and reserve (5′-A GCGGATAACAAT TTCACACAGG-3′) primers were used in combination with TDP2 reverse (5'-GAATTCTCATCACAGGATAAT-3') and TDP2(S101) forward or TDP2(G111) forward (5'-CATATGAGTCCGAGCGAAGATACC-3', 5'-CATATGGGCAGTATGTTTAGCCTGA-3', respectively) to generate the diagnostic PCR amplicon.

The subsequent amplification of recombinant baculovirus and the determination of viral titres followed the manufacturer's recommended protocol.

2.2.5 Expression and purification of full-length human PARP1

2x 0.5L Sf9 cell culture ($2x10^6$ cell/ml, at a viability greater than 95%) was infected with a P3 (passage 3) viral stock that expressed full length, human PARP1 (Dr Raquel Arribas / Dr Antony W. Oliver, University of Sussex), at a MOI (multiplicity of infection) of 2. Infected cultures were then incubated at 27°C and 150 rpm in a Multitron shaking incubator. After growth for 3 days, cells were harvested by centrifugation in an Avanti J-26XP centrifuge (Beckman Coulter) with JLA8.1 rotor at 1,000 rpm for 10 min at 4°C. The resulting cell pellet was stored at -80°C until required.

Purification of full-length human PARP1 essentially followed the protocol described by Giner at all (1992). The purified protein was concentrated to 4.7mg/ml (4.1μM) and then stored in aliquots at -80°C until required.

2.2.6 Expression and purification of TDP2

2.2.6.1 Cloning and expression trials of full length human TDP2

2.2.6.1.1 Cloning of full length human TDP2

A pUC57 plasmid containing the cDNA encoding full length human TDP2 (pUC57-hTDP2) was synthesised by GenScript, UK. The synthetic gene was flanked by *Nde*I and *Eco*RI restriction sites. 1μg of the pUC57-hTDP2 plasmid was cut with both *Nde*I and *Eco*RI (New England Biolabs, NEB) to release the DNA encoding TDP2. DNA migrating at the expected size was gel-extracted from a 1%, 1x TAE, and then ligated into the corresponding *Nde*I and *Eco*RI sites of both pTWO-E and pTHREE-E.

The appropriate recombinant expression constructs were identified by bacterial transformation and restriction digest of plasmid minipreps. The resultant constructs were termed pTWO-E/hTDP2 and pTHREE-E/hTDP2.

2.2.6.1.2 Small-scale expression trials

Experimental parameters are summarised in the following table (Table2.1):

Expression	
Construct	pTWO-E/hTDP2 and pTHREE-E/hTDP2
E. coli strain	BL21(DE3)
Growth media	LB
Antibiotic selection	50μg/ml Ampicillin
Growth parameters	37°C, 220 rpm in shaking incubator until OD_{600} =1
Induction of protein expression	Addition of 0.4mM IPTG, then adjustment of growth temperature (37°C for 3 hours, or 20°C overnight)
Lysis buffer	25mM HEPES-NaOH pH 7.5, 500mM NaCl, 0.5mM TCEP, 5% v/v glycerol

Table 2.1. Summary table of expression parameters for the full length human TDP2.

2.2.6.2 Cloning and expression trials of the catalytic domain of hTDP2 (hTDP2-CAT$^{101-362}$ and hTDP2-CAT$^{111-362}$)

To create expression vectors containing just the catalytic domain of hTDP2, the original synthetic hTDP2-pUC57 plasmid was used as a template for PCR amplification using the

follow primer sets: Forward [5'-CATATGAGTCCGAGCGAAGATACC-3'] and Reverse [5'-GAATTCTCATCACAGGATAAT-3'] amplified the region coding for amino acids Ser101-Leu362, and Forward [5'-CATATGGGCAGTATGTTTAGCCTGA-3'] and Reverse [5'-GAATTCTCATCACAGGATAAT-3'] amplified the region coding for amino acids Gly111-Leu362. Forward and reverse primers were designed to include flanking *Nde*I and *Eco*RI restriction sites, respectively.

PCR protocol:

2 min denaturation step at 95°C, 20 extension cycles (15 sec at 95°C, 15 sec at 58°C and 30 sec at 72°C). The final extension step was at 72°C for 210 sec. Amplicons of the expected size were gel-purified from 1% (w/v) agarose, 1x TAE gels and then ligated into both pTWO-E and pTHREE-E vectors, at the corresponding *Nde*I and *Eco*RI sites. The generated expression constructs were pTWO/hTDP2$^{101-362}$, pTWO-E/hTDP2$^{111-362}$ pTHREE-E/hTDP2$^{101-362}$, and pTHREE-E/hTDP2$^{111-362}$. Small-scale expression trials for these constructs were carried as per section (3.2.1.2).

2.2.6.3 Expression and purification of the mutant (C161S) catalytic domain of TDP2 (hTDP2-CAT^{C161S})

Expression and purification parameters are summarised in the following table (Table 2.2).

Expression	
Construct	pET28A/hTDP2-CAT^{C161S}
E. coli strain	B834(DE3)
Growth media	Turbo broth
Antibiotic selection	50 µg/ml Kanamycin
Growth parameters	37 °C, 220 rpm, until OD_{600} =2
Induction of protein expression	Addition of 0.4mM IPTG, then adjustment of growth temperature to 30°C, for 3.5 hours

Purification	
Lysis/washing buffer	50mM HEPES, pH7.5, 250mM NaCl, 10mM Imidazole and 0.5mM TCEP
Cell disruption	Sonication
1st Chromatography step	IMAC, gravity flow column
Resin	5ml TALON
Elution buffer	Batch elution; 50mM HEPES, pH7.5, 250mM NaCl, 0.3M Imidazole and 0.5mM TCEP
2nd Chromatography step	5ml HiTrap Heparin HP column (5ml) elution by linear salt gradient from 0% to 50% high salt buffer
Low salt buffer	50mM HEPES pH7.5, 150mM NaCl, 1mM TCEP
High salt buffer	50mM HEPES pH7.5, 2M NaCl, 1mM TCEP
3rd Chromatography step	HiLoad™ 16/600 Superdex™ 75pg (GE Healthcare)

Table continues on the next page

Gel filtration buffer	20mM HEPES pH7.5, 300mM NaCl and 0.5mM TCEP
Storage of protein	Concentrated to 7.3mg/ml and stored at -80°C.

Table 2.2. Summary table of expression and purification parameters for the catalytic domain of hTDP2. The His-tag was not removed during the purification procedure.

2.2.6.4 Expression and purification of Selenomethionine labelled hTDP2-CAT^{C161S}

Expression parameters are summarised in the following table (Table 2.3).

Expression	
E. coli strain	BL834 (DE3)
Growth media	Selenomethionine containing media (Section 2.2.6.4.1)
Antibiotic selection	50µg/ml Kanamycin
Growth parameters	37 °C, 220 rpm, until OD_{600} =1
Induction of protein expression	Adjustment of growth temperature to 30°C, then addition of 0.4mM IPTG, and induction for 3.5 hours.

Table 2.3. Summary table of expression parameters for the Selenomethionine labelled hTDP2-CAT^{C161S}. The purification protocol is identical to the procedure described for hTDP2-CAT^{C161S} (Section 2.2.6.3).

2.2.6.4.1 Preparation of Selenomethionine containing minimal media

21.6g of SelenoMet Medium Base (Molecular Dimensions) was dissolved in 1L ultrapure water and sterilized by autoclaving. 5.1g SelenoMet Nutrient Mix was dissolved in 50ml ultrapure water and sterilized by filtering. The solution was added to the minimal media alongside with 4ml Selenomethionine solution (250x), aseptically.

2.2.6.5 Expression and purification of the catalytic domain of mTDP2-CAT

Expression and purification parameters are summarised in the following table (Table 2.4).

Expression	
Construct	pET15b/mTDP2-CAT
E. coli strain	Rosetta 2(DE3)pLysS
Growth media	LB
Antibiotic selection	50µg/ml Ampicillin and 34µg/ml Chloramphenicol
Growth parameters	37°C, 220 rpm, until OD_{600} =1
Induction of protein expression	Addition of 0.4mM IPTG, then induced at 37°C for 4 hours

Purification	
Lysis/washing buffer:	50mM HEPES, pH7.5, 250mM NaCl, 10mM Imidazole, 1mM TCEP
Cell disruption	Sonication
Centrifuge step	50 min at 40000g at 4°C
1st Chromatography step:	IMAC, gravity flow column
Resin:	5ml TALON

Table continues on the next page

Elution buffer	Batch elution; 50mM HEPES, pH7.5, 250mM NaCl, 0.3M Imidazole and 0.5mM TCEP
Desalting step	HiPrep™ 26/10 desalting column
Desalting buffer	50mM HEPES pH7.5, 150mM NaCl, 1mM TCEP, SENP1 protease added for overnight digestion at 4°C
2nd Chromatography step:	5ml HiTrap™ Heparin HP column
Elution	Linear salt gradient, 0% to 50% high salt buffer
Low salt buffer:	50mM HEPES pH7.5, 150mM NaCl, 1mM TCEP
High salt buffer	50mM HEPES pH7.5, 2M NaCl, 1mM TCEP
3rd Chromatography step:	HiLoad™ 16/600 Superdex™ 75pg
Gel filtration buffer:	20mM HEPES pH7.5, 300mM NaCl and 0.5mM TCEP
Storage of protein	Concentrated to 9.6mg/ml and stored at -80°C.

Table 2.4. Summary table of expression and purification parameters for the catalytic domain of the murine TDP2 (mTDP2-CAT).

2.2.6.6 Expression and purification of the catalytic domain of the humanised mTDP2-CAT (m2hTDP2)

Expression and purification parameters are summarised in the following table (Table 2.5).

Expression	
Expression construct	pET15b/m2hTDP2-CAT
Expression strain	Rosetta 2(DE3)pLysS
Growth media	Turbo broth
Antibiotic selection	50µg/ml Ampicillin and 34µg/ml Chloramphenicol
Growth parameters	37°C, 220 rpm, until OD_{600} =2
Induction of protein expression	Addition of 0.4mM IPTG, then adjustment of growth temperature to 20°C, growth overnight

Purification	
Lysis/washing buffer:	50mM HEPES, pH7.5, 250mM NaCl, 10mM Imidazole, 1mM TCEP
Cell disruption	Sonication
Centrifuge step	50 min at 40000g at 4°C
1st Chromatography step:	IMAC, gravity flow column
Resin:	5ml TALON
Elution buffer	Batch elution; 50mM HEPES, pH7.5, 250mM NaCl, 0.3M Imidazole and 0.5mM TCEP
Desalting step	HiPrep 26/10 desalting column
Desalting buffer	50mM HEPES pH7.5, 150mM NaCl, 1mM TCEP, SENP1 protease added for overnight digestion at 4°C
2nd Chromatography step:	5ml HiTrap™ Heparin HP column
Elution	Linear salt gradient, 0% to 50% high salt buffer
Low salt buffer:	50mM MES, pH6.5, 150mM NaCl, 1mM TCEP
High salt buffer	50mM MES, pH6.5, 2M NaCl, 1mM TCEP

Table continues on the next page

3rd Chromatography step:	HiLoad™ 16/600 Superdex™ 75pg
Gel filtration buffer:	20mM MES, pH6.0, 300mM NaCl, 1mM TCEP
Storage of protein	Concentrated to 7.4mg/ml and stored at -80°C.

Table 2.5. Summary table of expression and purification parameters for the catalytic domain of the humanised mTDP2-CAT (m2hTDP2-CAT).

2.2.7 Expression and purification of the catalytic domain of the human tankyrase 1 (TNKS-1)

The expression and purification of the catalytic fragment of the human tankyrase 1 followed the protocol described in Lehtiö et al (2008).

2.2.8 Expression and purification of Human Sentrin-specific protease 1 catalytic domain (SENP1)

Expression and purification parameters are summarised in the following table (Table 2.6).

Expression	
Expression construct	pGEX-6P-1/SENP1 (residues 427 to 644)
Expression strain	Rosetta™ 2(DE3)pLysS
Growth media	LB
Antibiotic selection	100 μg/ml Carbenicillin and 34 μg/ml Chloramphenicol
Growth parameters	37°C, 220 rpm, until OD_{600} =2
Induction of protein expression	Addition of 0.4mM IPTG, then adjustment of growth temperature to 16°C, for overnight

Purification	
Lysis/washing buffer:	50mM HEPES pH 7.5, 500mM NaCl, 0.5mM TCEP, 10% v/v glycerol
Cell disruption	Sonication
Centrifuge step	50 min at 40000g at 4°C
1st Chromatography step:	Glutathione Sepharose 4 Fast Flow resin, gravity flow
Elution buffer	25mM MES pH 6.5, 50mM NaCl, 0.5mM TCEP, 5% v/v glycerol, 20mM Glutathione
2nd Chromatography step:	5ml HiTrap™ SP HP column
Elution	Linear salt gradient, 0% to 100% high salt buffer
Low salt buffer:	25mM HEPES pH7.5, 50mM NaCl, 5% v/v glycerol, 0.5mM TCEP
High salt buffer	25mM HEPES pH7.5, 1M NaCl, 5% v/v glycerol, 0.5mM TCEP
3rd Chromatography step:	HiLoad™ 26/60 Superdex™ 200pg
Gel filtration buffer:	25mM HEPES pH 7.0, 150mM NaCl, 0.5mM TCEP and 5% v/v glycerol
Storage of protein	Concentrated to 2mg/ml and stored at -80°C in 50% glycerol.

Table 2.6. Summary table of expression and purification parameters for the catalytic domain of SENP1.

2.2.9 Expression and purification of MACROD1

Expression and purification parameters are summarised in the following table (Table 2.7).

Expression	
Expression construct	pET16b/MACROD1 (the vector was a kind gift of Dr Gabrielle Grundy)
Expression strain	BL21(DE3)pLysS
Growth media	LB
Antibiotic selection	50µg/ml Ampicillin
Growth parameters	37 °C, 220 rpm, until OD_{600} =1
Induction of protein expression	Adjustment of growth temperature to 16°C, addition of 0.4 mM IPTG, induction for overnight

Purification	
Lysis/washing buffer:	50mM Tris-HCl pH7.5, 150mM NaCl, 10mM Imidazole and 0.5mM TCEP
Cell disruption	Sonication
Centrifuge step	50 min at 40000g at 4°C
1st Chromatography step:	IMAC, gravity flow column
Resin:	5ml TALON
Elution buffer	Batch elution; 50mM Tris-HCl pH7.5, 150mM NaCl, 300mM Imidazole and 0.5mM TCEP
2nd Chromatography step:	HiLoad™ 16/600 Superdex™ 75pg
Gel filtration buffer:	20mM Tris-HCl, pH7.5, 300mM NaCl, 0.5mM TCEP
Storage of protein	Concentrated to 0.4mg/ml and stored at -80°C.

Table 2.7. Summary table of expression and purification parameters for full length MACROD1.

2.2.10 Cloning and expression of XRCC1$^{161-406}$

2.2.10.1 Cloning of XRCC1$^{161-406}$ wild type and R335A/K369A mutant

To create the extended XRCC1-BRCT1 expression constructs, the parental plasmids pCD2EXHR^{WT} and pCD2EXH$^{R335A/K369A}$ (encoding human XRCC1-His WT and R335A/K369A, respectively) were used as templates. The ORF encoding XRCC1 residues 161-406 were amplified by PCR using oligonucleotides 5'-CATATGCCGTCCCAGAAGGTGAC-3' (forward primer) and 5'-GAATTCTCATCATGGCCCTGCCATGAGG-3' (reverse primer) (from Eurogentec). The PCR protocol was as follows: 2 min denaturation step at 95°C, 20 extension cycles of: denaturation for 15 sec at 95°C, annealing for 15 sec at 56°C and elongation for 30 sec at 72°C, followed by a final extension step at 72°C for 210 sec, before cooling to 4°C.

PCR products were loaded onto 1% Agarose gel supplemented with ethidium bromide and run for 30 min at 140V at constant current of 400mA. The main band was cut out from the

gel and purified with an illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and then captured by cloning into pCR4 Blunt-TOPOvector (Thermo Fisher Scientific) (the TOPO cloning methodology is fully described in 2.1.6). The plasmids were transformed into DH5α *E. coli* competent cells and were streaked onto LB plates supplemented with 50µg/ml kanamycin and incubated overnight at 37°C. The following day single colonies were picked and inoculated into 5ml LB media with 50µg/ml kanamycin and were incubated overnight at 37°C. Next day plasmids were extracted using QIAprep® Spin Miniprep kit (Qiagen). Plasmids were subjected to double digest using the restriction enzymes *Nde*I and *Eco*RI to verify clones with insert. Positive clones then were sent for sequencing for further verification. The insert was excised by *Nde*I-*Eco*RI (both enzyme from New England Biolabs, NEB) restriction digest, then loaded onto 1% agarose gel. Following gel electrophoresis the correct band was gel extracted and ligated into an *Nde*I-*Eco*RI double digested, Calf Intestinal Phosphatase (CIP, from NEB) treated pTWO-E vector. 5µl of the ligation reaction was transformed into DH5α *E. coli* competent cells. Cells were streaked onto LB plates, supplemented with 50µl ampicillin and incubated at 37°C for overnight. Next day single colonies were inoculated into 5ml LB with 50µl ampicillin and incubated for overnight at 37°C. The following day plasmids were extracted and subjected to diagnostic digest. Positive clones were sent for sequencing.

2.2.10.2 Expression and purification of XRCC1$^{161\text{-}406\ WT}$ and XRCC1$^{161\text{-}406\ RK}$

Expression and purification parameters are summarised in the following table (Table 2.8).

Expression	
Expression construct	pTWO-E/ XRCC1$^{161\text{-}406\ WT}$ and pTWO-E/ XRCC1$^{161\text{-}406\ RK}$
Expression strain	Rosetta™2 (DE3)pLysS
Growth media	LB
Antibiotic selection	50µg/ml Ampicillin and 34µg/ml Chloramphenicol
Growth parameters	37 °C, 220 rpm, until OD_{600} =1
Induction of protein expression	Adjustment of growth temperature to 30°C, addition of 0.4 mM IPTG, induction for 4 hours

Purification	
Lysis/washing buffer:	50mM HEPES, pH7.5, 250mM NaCl, 10mM Imidazole, 1mM TCEP
Cell disruption	Sonication
Lysate clarified by	Centrifugation for 50 min at 40000g at 4°C
1st Chromatography step:	IMAC, gravity flow column
Resin:	5ml TALON
Elution buffer	Batch elution; 50mM HEPES, pH7.5, 250mM NaCl, 0.3M Imidazole and 0.5mM TCEP

Table continues on the next page

2nd Chromatography step:	5ml HiTrap™ Heparin HP column
Elution	Linear salt gradient, 0% to 50% high salt buffer
Low salt buffer:	50mM HEPES pH7.5, 150mM NaCl, 1mM TCEP
High salt buffer	50mM HEPES pH7.5, 2M NaCl, 1mM TCEP
Storage of protein	Concentrated to 0.3mg/ml and stored at -80°C.

Table 2.8. Summary table of expression and purification parameters XRCC1$^{161-406}$ truncated proteins constructs.

2.3 TDP2 activity assays

2.3.1 Labelling protocol

Oligonucleotide A: 5'-Y-P-AATTCTTCTCTTTCCAGGGCTATGT-3; where Y = 5'-phosphotyrosine link (Midland Certified Reagent Company, USA)

Oligonucleotide B: 5'-AGACATAGCCCTGGAAAGAGAAG-3' (Sigma-Aldrich).

For annealing 2µl of 10µM 5´Y oligonucleotide A and 4µl of 10µM oligonucleotide B were mixed together, in a final volume of 30µl in ultrapure water. Oligonucleotide B was added in excess to maximize the complete annealing of oligonucleotide A. The annealing reaction was completed in an *MWG Biotech Primus 25* thermal cycler and the following program used: 95°C for 10min, 70°C for 10, 45°C for 10min, 20°C for 10min, and finally 4°C for 10min; ramp: 5°C/min.

The annealed DNA duplex was then labelled using Klenow Fragment (3'→5' exo$^-$; New England BioLabs) in the presence of 2 µl of 10mM ddTTP (GE Healthcare) and 2µl α-^{32}P dCTP (Perkin Elmer) in a reaction volume of 40µl. The reaction was incubated for 1 hour at 37°C, in a heating block, and then the DNA duplex purified away from any unincorporated radiolabel using an Illustra MicroSpin G25 column (GE Healthcare) at 3000 rpm for 2 min at room temperature and then stored at -20°C until required. The final concentration of DNA duplex was approximately 500nM.

2.3.2 Enzymatic activity assay

hTDP2-CAT^{C161S}, mTDP2-CAT and m2hTDP2-CAT were mixed with reaction buffer (50 mM Tris-HCl, pH 7.5, 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 100 µg/ml BSA); at a final DMSO concentration of 1% v/v. Reactions were set up with or without the inhibitory compound, in a compound: protein molar ratio of 50:1, with a volume of 16.2µl. To start the reaction 1.8µl labelled substrate DNA was included, making a final reaction volume of

18µl. The final concentrations of hTDP2-CAT, mTDP2-CAT and m2hTDP2-CAT were 1nM, 3nM and 1nM, respectively; mTDP2-CAT may have been not enzymatically as active as the other two enzymes. After incubation at 37°C, and withdrawal of 6µl samples taken at 2, 5 and 10 min time points, the reaction was stopped by the addition of 1x loading buffer containing formamide. Samples were loaded onto an SDS-Urea gel [20% Acrylamide:Bis-Acrylamide 19:1 (Thermo Fisher Scientific), 36% w/v Urea (Thermo Fisher Scientific) and 1x TBE buffer] and run in a V15·17 vertical gel electrophoresis apparatus (Thermo Fisher Scientific) at a constant power of 20W for 80 min. When the run was completed, the gel was placed into a BAS-MS 2025 Imaging Plate (Fuji, UK) for 15 min. The plate was then scanned by a Fujifilm FLA-5100 imaging system in standard reading mode with a 635nm laser setting.

2.4 Crystallisation trials with purified recombinant proteins

2.4.1 Crystallization of TDP2

For initial screening MRC 2 well crystallisation plates (Swissci, Hampton Research, USA) were used in a sitting drop vapour diffusion set up. The plate is a two drop chamber format with 96 wells. Buffers and protein solutions were dispensed using the Phoenix robot system. The initial trials employed crystallisation screens from Hampton research.

When additional (or secondary) screening was required, hanging drop vapour diffusion method was used employing 24-well VDX Plates (with sealant, Hampton Research). All crystallisation trials were undertaken at 20°C.

2.4.1.1 Crystallisation of the unliganded (apo-form) of hTDP2-CAT^{C161S}

Hanging drop 24-well VDX Plates: 1µl 7.3mg/ml hTDP2-CAT^{C161S} mixed with 1µl mother liquor: 0.45M NaCl, 5mM $MgCl_2$, 100 mM bis-tris propane pH7.0 and 1.5% v/v trimethylamine *N*-oxide – and equilibrated with 1 ml of mother liquor. After ~5 days, crystals were harvested and frozen in liquid nitrogen in the presence of 40% w/v sucrose as a cryoprotectant.

2.4.1.2 Crystallisation of the hTDP2-CAT^{C161S} in complex with inhibitors

Compound C14297 (22.5mM in 90% v/v DMSO and 100mM HEPES-NaOH pH7.5) was added to the protein (final compound: protein molar ratio was 2.7:1, with a final DMSO concentration of 2.7 % v/v, TDP2 concentration was 7.2mg/ml) and then incubated on ice for 30mins.

Sitting drop MRC 2 well plates: 0.3µl protein-compound solution was mixed with 0.3µl crystallization buffer and equilibrated with 50µl mother liquor. Index, PEG/Ion and SaltRx HT crystallisation screens were set up (Hampton Research). After ~7 days, crystals from condition C9 of the SaltRx HT (1.2M DL Malic acid pH 7.0, 0.1M bis-tris propane pH7.0, Hampton Research) were harvested and then plunge-frozen in liquid nitrogen in the presence of the crystallisation buffer with 30% v/v glycerol added as a cryoprotectant.

2.4.1.3 Crystallisation of Selenomethionine-labelled hTDP2-CAT^{C161S}

Hanging drop 24-well VDX Plates: 1µl 7.3mg/ml hTDP2-CAT^{C161S} mixed with 1µl mother liquor: 0.45M NaCl, 5mM $MgCl_2$, 100 mM bis-tris propane pH7.0 and 1.5% v/v Trimethylamine *N*-oxide – and equilibrated with 1 ml of mother liquor. After ~5 days, crystals were harvested and frozen in liquid nitrogen in the presence of 40% w/v sucrose as a cryoprotectant.

2.4.1.4 Crystallisation of m2hTDP2-CAT with C14297

Compound C14297 (22.5mM in 90% v/v DMSO and 100mM HEPES-NaOH pH7.5) was added to the protein (final compound: protein molar ratio was 2.7:1, with a final DMSO concentration of 2.7 % v/v, TDP2 concentration was 7.2mg/ml) and then incubated on ice for 30mins.

Sitting drop MRC 2 well plates: 0.2µl protein-compound solution was mixed with 0.2µl crystallization buffer and equilibrated with 50µl mother liquor. Index, PEG/Ion and SaltRx HT crystallisation screens were set up (Hampton Research). After ~5 days, crystals from condition G11 of the PACT *premier*^{TM} HT (0.2M Sodium Citrate, 0.1M bis-tris propane pH7.5, 20% w/v PEG 3,350, Hampton Research) were harvested and then plunge-frozen in liquid nitrogen in the presence of the crystallisation buffer with 8% glycerol and 5% glucose added as a cryoprotectant.

2.4.1.5 Crystallisation of m2hTDP-CAT with LEP-0259

Compound LEP-0259 (22.5mM in 90% v/v DMSO and 100mM HEPES-NaOH pH7.5) was added to the protein (final compound: protein molar ratio was 2.7:1, with a final DMSO concentration of 2.7 % v/v, TDP2 concentration was 7.2mg/ml) and then incubated on ice for 30mins.

Sitting drop MRC 2 well plates: 0.2µl protein-compound solution was mixed with 0.2µl crystallization buffer and equilibrated with 50µl mother liquor. Index, PEG/Ion and SaltRx

HT crystallisation screens were set up (Hampton Research). After ~5 days, crystals from condition F9 of the PACT *premier*™ HT (0.2M Potassium sodium tartrate tetrahydrate, 0.1M bis-tris propane pH6.5, 20% w/v PEG 3350, Hampton Research) were harvested and then plunge-frozen in liquid nitrogen in the presence of the crystallisation buffer with 8% glycerol and 5% glucose added as a cryoprotectant.

2.5 *In vitro* binding assays

2.5.1 *In vitro* binding assay, Slot blot

2.5.1.1 Slot Blot, XRCC1$^{161-406}$

50µl of recombinant XRCC1$^{161-406\ WT}$ or XRCC1$^{161-406\ RK}$ at 100nM was first diluted in binding buffer (20mM Tris-HCl pH7.5, 100mM NaCl) and then transferred onto Amersham Protran nitrocellulose membrane (GE Healthcare), using a PR 648 Slot Blot Manifold, 48-well device (GE Healthcare) by applying vacuum. Following transfer, the membrane was blocking for 30 minutes at room temperature in 5% w/v milk powder (Marvel, UK) dissolved in binding buffer. The membranes were subsequently incubated with either mock-ribosylated or auto-ribosylated recombinant PARP1 for 30 min at room temperature.

Ribosylation reactions were as follows: 100nM PARP1 enzyme was mixed either with zero or 10µM NAD^+ (Sigma-Aldrich), 20nM single-stranded oligodeoxyribonucleotide (5'-CATATGCCGGAGATCCGCCTCC-3', Eurogentec) and ultrapure water to a final volume of 500µl. The reaction was allowed to proceed at room temperature for 30 min.

After incubation with PARP1, the membrane was washed twice with 2% w/v milk-binding buffer, and then incubated with either 1:300 dilution mouse anti-PARP1 (#A6.4.12, Bio-Rad) or 1:1000 mouse anti-Poly(ADP-ribose) (#10H, Enzo Life Sciences, Exeter, UK) monoclonal antibodies in 2% w/v milk-binding buffer overnight at 4°C. The next morning the membranes were washed 3 times with 2% w/v milk-binding buffer and then incubated with 1:5000 dilution HRP-conjugated rabbit anti-mouse IgG (Dako, Ely, UK) for 1h at room temperature. The membrane was washed again with milk-binding buffer, and then placed into Amersham ECL Western Blotting Detection Reagent (GE Healthcare). Amersham Hyperfilm Blue (GE Healthcare) was used to detect the resultant chemiluminescent signal.

2.5.1.2 Slot Blot, PAR blotted

A dilution series of PAR polymer (Trevigen, Abingdon, UK) was prepared in phosphate-buffered saline buffer (PBS: 137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4, and 1.47mM KH_2PO_4), and then slot-blotted (as before) onto an Amersham Hybond-XL nylon membrane (GE Healthcare). The membrane was allowed to dry at room temperature for 10 min after vacuum transfer, and then cross-linked by exposure to UV light for 25 sec

using UV Stratalinker 2400 (Stratagene, USA) ultraviolet cross-linker device. After blocking for 30 min in 5% w/v milk-binding buffer at room temperature the membrane was incubated with either 100nM XRCC1$^{161\text{-}406\ WT}$ or XRCC1$^{161\text{-}406\ RK}$ protein diluted in binding buffer and then incubated at room temperature for a period of 30 min. After three wash steps using 2% w/v milk-binding buffer, the membrane was incubated with 1:3000 mouse monoclonal anti-polyhistidine antibody (#HIS-1, Sigma-Aldrich) for overnight at 4°C. The next morning after three washes with 2% w/v milk-binding buffer the membrane was probed as before with 1:5000 HRP-conjugated rabbit anti-mouse IgG (Dako) for 1h at room temperature. The membrane was then washed as per the previous protocol (Section 2.5.1.1), and any retained protein detected as before. As a control, a membrane containing only dilution series of PAR was probed for anti-Poly(ADP-ribose)(10H) (1: 1000, Enzo Life Sciences) antibody to measure the level of PAR signal emerging from each dilution.

2.5.2 Plate binding assays

For plate-binding assays, the wells of a flat-bottomed 96 well PS-microplate (Greiner, Germany) were filled with 50 µl of either recombinant histone H1 (protein stock was a kind gift of Dr Gabrielle Grundy, University of Sussex), PARP1, BSA (Sigma-Aldrich) or Calf Thymus Histone type II (Sigma-Aldrich) at 0.1 mg/ml in PBS was absorbed onto plates at 4°C overnight. The next day the wells were rinsed four times with 0.2 ml PBS containing 0.1% v/v Triton X100 (Sigma-Aldrich).

2.5.2.1 Poly(ADP-ribose) binding assays

Adsorbed histone H1, PARP1 and BSA were mock-ribosylated (in the absence of NAD^+) or trans-ribosylated in the presence of the indicated concentration of NAD^+ (Sigma-Aldrich), in 50 µl of PARP1 reaction buffer (50 mM Tris-HCl pH 8, 0.8mM $MgCl_2$, 1% v/v glycerol and 1.5 mM DTT) containing 40nM single-stranded oligodeoxyribonucleotide (see section 2.5.2.1) and 5nM PARP1 (recombinant, human, full length) at room temperature for 30min. After four successive rinses with 50 µl of PBS containing 0.1% v/v Tween 20, 50 µl of either XRCC1$^{161\text{-}406\ WT}$ or XRCC1$^{161\text{-}406\ RK}$ (diluted to 25nM in 20 mM Tris pH 7.5, 130mM NaCl) was added to the wells and then incubated on ice for 30min. The wells were again rinsed four times as above, and then incubated with 50µl of 20 mM Tris pH7.5, 130nM NaCl containing 1:3000 mouse monoclonal anti-polyhistidine antibody (Sigma-Aldrich). The wells were rinsed again (4x) followed by incubation with 50µl buffer containing 1:500 HRP-conjugated rabbit anti-mouse IgG (Dako) for 30min on ice. After a

final wash with 50 µl of PBS containing 1% v/v Tween 20, 50µl of TACS Sapphire reagent (Trevigen) was added to the wells and incubated for 15min in the dark. The colorimetric reaction was stopped by adding 50µl of 0.2 M HCl, and the absorbance was read at 450nm in a POLARstar Omega (BMG Labtech, Germany) microplate reader.

2.5.2.2 Auto- and trans-ribosylated PARP1 competition assay

Before XRCC1 proteins were added to the wells, they were pre-incubated with mock-ribosylated and auto-ribosylated PARP1 (with the indicated concentration of NAD^+ for 30 min at 4°C. The washing and detections steps were the same as described in Section 2.5.2.1.

2.5.2.3 Mono(ADP-ribose) and poly(ADP-ribose) competition assays

Before XRCC1 proteins was added to the wells, they were pre-incubated with mono(ADP-ribose) (Sigma-Aldrich) or poly (ADP-ribose) (Trevigen) competitor at the concentrations indicated for 30 min at 4°C. The washing and detections steps were the same as described in Section 2.5.2.1. The same competition assay was also conducted in the presence of purified PAR fractions with distinct unit lengths in a sequential dilution series (15µM, 1.5µM, 0.15µM and 15nM).

2.5.2.4 Plate binding assays employing biotinylated PAR polymer

Adsorbed Calf thymus histone was either mock-ribosylated or trans-ribosylated with biotinylated NAD^+ in the presence of PARP1. For mono-ribosylation reactions 100nM recombinant PARP3 in PARP reaction buffer (50mM Tris-HCl pH8, 0.8mM $MgCl_2$, 1% glycerol and 1.5mM DTT) was used, and was incubated at room temperature for 30 min.

PAR polymer was digested with 1ng/ml PARG enzyme (Trevigen) in 50µl reaction buffer (20mM KPO4 (pH 7.2), 50mM KCl, 0.1 mg/ml BSA, 0.1% Triton® X-100, 1mM DTT, 5% Glycerol) at room temperature for 30 min.

Mono-ribosyl moieties from the acceptor protein were removed by digestion with 15µg/ml recombinant MACROD1 in PARP reaction buffer at room temperature for 30 min.

Biotinylated PAR polymer and mono(ADP-ribose) were detected by HRP-conjugated streptavidin antibody (1:10,000 dilution in 20mM Tris-HCl pH7.5, 130mM NaCl; Thermo Fisher Scientific).

2.5.3 Preparation of homogenous PAR polymer

2.5.3.1 Preparation of dihydroxyboryl groups coupled to BioRex 70 beads

The full protocol is described in (Bredehorst et al. 1981). In short, 50g Bio-Rex resin (100-200 mesh, Bio-Rad) was incubated with 5g aminophenyl-boronic acid (Sigma-Aldrich) and 5g N-ethyl-N'-(3-diethylaminopropyl)-carbodiimide (Sigma-Aldrich) overnight at pH 5.0 at 20°C. The next day the sample was washed with water 3 times, then with 0.1M Na acetate, and 1M NaCl buffer (pH 4.5), then with 0.1M $NaHCO_3$, and 1M NaCl buffer (pH 9.0), and again with water. The resin was finally washed with 6M guanidine HCl, and then 0.5M MES buffer (pH 6.0) and stored in the same buffer.

2.5.3.2 Preparation of large scale homogenous PAR polymer

The purification protocol is based on (Tan et al. 2012) with some minor alterations. The PARylation reaction was as follows: 1mg/ml calf thymus histones (Sigma-Aldrich) in PARP reaction buffer (50mM Tris-HCl pH 8, 0.8mM $MgCl_2$, 1% v/v glycerol and 0.5mM DTT), 20mM NAD^+ (Sigma-Aldrich) were mixed with 1mg/ml TNKS1 enzyme. The total reaction volume was 1ml and the reaction was incubated at room temperature for 1 hour. The reaction was stopped by adding an equal volume of 20% v/v ice-cold trichloroacetic acid, then incubation on ice for 15 min. The precipitated ribosylated protein was pelleted by centrifugation at top speed at 4°C. After decanting the supernatant, the pellet was dissolved in 100µl 1M KOH/50mM EDTA and was incubated for 60 min at 60°C with regular vortexing. After the incubation period, 900µl of AAGE9 buffer (250mM ammonium acetate, 6M guanidine HCl, 10mM EDTA, pH 9.0) was added and the sample was loaded onto 1ml dihydroxyboryl Bio-Rex resin pre-equilibrated with AAGE9 buffer. The resin was washed (in a gravity column format) with 10ml AAGE9 buffer, 20ml 1M ammonium acetate (pH 9.0) buffer and eluted with 6ml water in 1ml fractions. Successive fractions were analysed by UV spectroscopy, using a NanoDrop2000 (Thermo Fisher Scientific) at a wavelength of 258nm. Fractions containing bulk PAR were pooled and then loaded onto a 1ml MonoQ 5/50 chromatography column (GE Healthcare). The column was first washed with Buffer A (25mM Tris-HCl, pH9.0) to remove any unbound material. Bound PAR was eluted by the application of the following linear gradient series from Buffer A to Buffer B (25mM Tris-HCl, pH 9.0, 1M NaCl): 0% to 15% B over 5 column volumes (cv), then 15% to 40% B over 130 cv, followed by 40% to 45% B over 80cv, and a final step from 45% to 100% B over 3cv. Fractioned peaks were dried in a Savant

DNA120 SpeedVacTM concentrator (Thermo Fisher Scientific) and then stored at -20°C until required. Successive runs of the MonoQ column were required in order to generate a sufficient quantity of fractionated PAR. Fractions containing PAR, with the same elution volume, were pooled together and re-dissolved in water to final volume of 1ml, which was then loaded onto a PD MidiTrap G-10 column (GE Healthcare) pre-equilibrated with water. After elution from the column (following the manufacturer's protocol) the resulting sample was again dried in the Savant concentrator before being stored at -20°C.

2.5.3.3 Acrylamide gel electrophoresis of purified PAR fractions

Purified PAR fractions were adjusted to a final concentration of 0.3µM and then diluted in loading buffer (40% w/v urea, 4mM EDTA, 0.02% w/v Bromophenol blue, and 0.02% w/v Xylene cyanol) to a final volume of 15µl and then loaded onto a 20% v/v polyacrylamide gel (acrylamide: bis-acrylamide ratio 19:1, Thermo Fisher Scientific) containing 1x TBE buffer. The gel was run at a constant power of 15W until the dye front migrated approximately 50% of the gel; after which the gel was fixed in a 50% v/v ethanol and 5% v/v acetic acid solution for 2 hours. After fixing, the gel was washed with 4 consecutive 15 min washes in ultrapure water. The gel was then stained with a Pierce Colour Silver Stain Kit (Thermo Fisher Scientific) following the manufacturer's protocol. Briefly: 'working solution' was added to the gel then incubated for 30 min, it was then rinsed briefly in ultrapure water, before 'reducer solution' was added and further incubation for 2 minutes. Next, the gel was incubated in 'stabiliser solution' for 30 min. The stained gels were then sealed in plastic bags and scanned using a scanning device (Epson) connected to a PC. All staining steps were carried out at room temperature.

2.6 *In vitro* biophysical assays

2.6.1 Thermal denaturation assay

For thermal denaturation, samples containing protein at a given concentration and 5 x SYPRO Orange (diluted from a 5000 x stock supplied in DMSO; catalogue number S5692, Sigma-Aldrich) were prepared in sample buffer [50mM HEPES pH 7.5, 300mM NaCl, 0.5mM TCEP]. Denaturation curves were monitored in 96-well PCR plates in a Roche LightCycler 480 II, using 465 and 580 nm filters for excitation and emission wavelengths, respectively. The program was as follows: 1 min at 20°C and then the plate was heated at a continuous increment of 0.03°C/s to 85°C. Temperature midpoints (T_m) for each folded

to unfolded transition were determined by either calculating a first-order derivative of the observed melting profile (built-in analysis tool supplied with the LightCycler) or by non-linear regression fitting of a modified Boltzmann model to normalized data in Prism5 (GraphPad Software).

$$Y = (a_n X + b_n) + \frac{(a_d X + b_d) - (a_n X + b_n)}{1 + e^{\frac{T_m - X}{m}}}$$

Where: a_n and a_d are the slopes, b_n and b_d the y-intercepts, of the native and denatured baselines respectively. T_m is the melting temperature, and m a slope factor.

2.6.2 Circular dichroism

For circular dichroism, spectra were measured between the wavelengths 198nm and 280nm, at 20 ˚C, in a JASCO J-715 spectropolarimeter attached to a JASCO PTC-384W temperature control system. Protein samples in 10mM HEPES, pH7.5, 300mM NaCl and 0.5mM TCEP were concentrated to 1.55mg/ml and put into a Type 20 demountable cell (path length 0.1mm, Starna Scientific). CD Spectra represent the average of 10 individual scans, where the absorbance from a buffer control has been subtracted.

2.7 Cell biology techniques

2.7.1 Maintaining cell cultures

Cells were maintained as a monolayer culture in either 75cm^2 or 175cm^2 flasks (T75, T175, Corning) grown in a NuAire Microbiological CO_2 incubator. Cells were provide with α-MEM media (Gibco) supplemented with 10% v/v Foetal Calf Serum (FCS), 1% v/v L-Glutamine (Gibco) and 1% v/v Penicillin-Streptomycin (10,000 U/mL) (Gibco). Cell manipulation and aseptic techniques were all performed in a NuAire Safety cabinet. Cell counts were determined with a Marenfield counting chamber (haemocytometer).

Stocks of cells were prepared in media containing 10% v/v DMSO at 0.5x10^6 cell count and stored in TruCool cryogenic vials (Biocision). Vials first were cooled down in a Mr Frosty Cryo 1°C Freezing Container (Thermo Fisher Scientific) and then transferred and stored in liquid N_2.

2.7.2 XRCC1 pull down assays

Stably transfected EM9 cells [a derivative of the Chinese hamster ovary (CHO) cell line, which is $XRCC1^{-/-}$] expressing full length His_6-tagged XRCC1, either wild-type (EM9 XH) or phosphopeptide-binding site mutant (EM9 R335A/K369A), were grown in T75 flasks in α-MEM media (Gibco) supplemented with 10% v/v Foetal Calf Serum (FCS), 1% v/v L-Glutamine (Gibco) and 1% v/v Penicillin-Streptomycin (10,000 U/mL) (Gibco). After reaching confluence, cells were washed with PBS and then trypsinized (0.5% w/v trypsin-PBS) to detach them from the flask and counted using a haemocytometer. $1x10^7$ cells were then transferred to a 15ml Falcon tube (Thermo Fisher Scientific) and were centrifuged at 1500 rpm for 5min at room temperature in a Hettich Rotina 38R centrifuge with a 4 place swing out rotor (A1724) (DJB Labcare, Newport Pagnell, UK). After aspirating the media the cell pellet was resuspended in 5ml fresh media and centrifuged as before. The media was once again aspirated, and the cell pellet re-suspended in ice cold PBS and it was centrifuged at 1500 rpm for 5min at 4°C. The cell pellet was resuspended in 1ml PBS and was transferred to a 1.5ml Eppendorf tube.

Cells were harvested by centrifugation at 1500 rpm at 4°C for 5 minutes in a Hettich Rotina 38R centrifuge with a 4 place swing out rotor (A1724) (DJB Labcare). Cells were washed once in 5 ml of fresh growth media, and once with 5ml ice-cold PBS, with centrifugation between each wash to harvest cells. The washed cell pellet was resuspended in 1ml PBS and transferred to a 1.5 ml Eppendorf tube. Again, cells were harvested by centrifugation (13,000 rpm at 4°C, 1 min, in a benchtop microcentrifuge) and then either snap-frozen in liquid N2 and stored at 80°C or processed immediately.

For pull down experiments, the following buffer was used: 25mM HEPES pH 7.8, 150mM NaCl, 10% v/v glycerol, 0.5% v/v Triton-X, 1/100 Protease inhibitor cocktail (Sigma-Aldrich) and 1/100 Phosphatase inhibitor cocktail 3 (Sigma-Aldrich) were prepared. Lysis/Washing buffer was comprised of the same buffer but supplemented with 1mM DTT and 25mM imidazole. The Elution buffer is comprised of the same buffer, but supplemented with 1mM DTT and 250mM imidazole.

Lysis buffer was added to the cell pellet in order to adjust the cell count to $400\mu l/5x10^6$ cells, and was then placed on ice for 20 min. The cells were disrupted by sonication for 2x 5 seconds at 20% power using a small parallel probe (Sonics Vibra-Cell, VWR), and cell debris and insoluble material removed by centrifugation at 13000 rpm at 4°C for 20 min

in a benchtop refrigerated centrifuge. 300 µl of the supernatant was added to 100µl pre-equilibrated Ni-agarose resin slurry (50µl resin and 50µl wash buffer) and then incubated with mixing on a rotating wheel at 4°C for 30 min. The resin was then pelleted by centrifugation at 3000 rpm for 1 min in benchtop refrigerated centrifuge. The flow-through was retained (unbound) and the beads were washed with five successive applications of 500 µl of wash buffer, pelleting the resin after each wash by centrifugation as before; the final wash was retained (wash). 200µl elution buffer was added to the resin and after resuspension incubated on ice for 5min. The resin was again pelleted by centrifugation and the supernatant (retained) was retained for further analysis. Finally, 125µl Laemmli buffer was added to the resin and was heated to 100°C for 5 min (beads). Laemmli buffer was added to the other samples (unbound, wash, retained) and heated to 100°C for 2min. 20µl of each sample were analysed by SDS-PAGE.

Western blot

After separation by SDS-PAGE, proteins were transferred to Amersham Protran nitrocellulose membrane (GE Healthcare) using acrylamide gel electrophoretic apparatus at 100V for 90 min, in TOWBIN transfer buffer (25mM Tris-Base pH 8.3, 192mM glycine and 10% v/v methanol).

After transfer, the membrane was blocked in 5% milk-PBS for 1 hour at room temperature. After blocking, the membrane was incubated with 1:3000 dilution of mouse monoclonal anti-PARP1 (Bio-Rad) antibody, diluted in 1% milk-PBS, overnight at 4°C. The next morning the membrane was washed 3 times with 1% milk-PBS and then probed with 1:5000 dilution of HRP-conjugated rabbit anti-mouse IgG (Dako) antibody for 1h at room temperature. The membrane was washed 4 times with PBs, and then soaked in Amersham ECL Western Blotting Detection Reagent (GE Healthcare). The resultant chemiluminescent signal was detected on Amersham Hyperfilm Blue film (GE Healthcare).

In order to re-probe for XRCC1, the membrane was stripped to remove the antibodies. Membranes were incubated with 4 successive applications of stripping buffer [40mM Tris-HCl pH 6.8, 0.6% v/v β-mercaptoethanol, 2% v/v SDS] for 10 min at 50°C, washed with ultrapure water, and 3 successive applications of Tris-buffered Saline (TBS) buffer with 0.1% v/v Tween 20 (Sigma-Aldrich) added (TBS-T).

Membranes were blocked in 5% milk TBS-T, then the presence of full length XRCC1 was detected with a 1:5000 dilution of rabbit anti-phospho XRCC1 polyclonal antibody (pS485/pT488) (A300-231A, Bethyl Laboratories) in 5% milk-TBST, or XRCC1$^{161-406}$ constructs with a 1:500 dilution of mouse anti-Myc monoclonal antibody (9B11, Cell Signalling Technology) in 5% milk-TBST. After successive washes to remove any unbound antibody (described above), either a 1:5000 dilution of anti-rabbit or anti-mouse HRP-conjugated IgG antibody (Dako) was used as secondary antibody. As before, detection of chemiluminescence was by film.

2.7.3 Clonogenic survival in U2OS cells using small molecule inhibitors of hTDP2 in the presence of etoposide (VP16)

The osteosarcoma cell line U2OS was maintained as a monolayer culture in 75cm^2 flasks (T75, Corning) in modified Eagle's medium (α-MEM) media (Gibco) supplemented with 10% v/v Foetal Calf Serum (FCS), 1% v/v L-Glutamine (Gibco) and 1% v/v Penicillin-Streptomycin (10,000 U/mL) (Gibco). Cell culture was maintained in a NuAire Microbiological CO_2 incubator at 37°C. After reaching approximately 90% confluence, the cells were washed with PBS and then trypsinised (as per EM9 cells, Section 2.7.2) to detach them from the flask, and then counted using a haemocytometer.

For 0, 200 and 400nM etoposide treatments, 500 cells were seeded; for 400, 600 and 800nM treatments 5000 cells were plated onto Petri culture dishes (Corning) in 10ml media. Dishes were incubated at 37°C for 4 hours, to allow the cells to adhere to the bottom of the dishes. Half of the dishes were treated with 5μM hTDP2 inhibitor in the presence of 0.5% v/v DMSO and the other half treated with 0.5% DMSO alone (as a vehicle control), and then incubated for 4 hours at 37°C in a NuAire Microbiological CO_2 incubator.

After the incubation period, the plates were treated with 0, 200, 400, 600 and 800nM etoposide (VP16) and was incubated for a further 4 days at 37°C, after which the plates were washed with PBS and fresh media was added. The dishes were then incubated for an further 10-12 days and checked for colony formation. Dishes were then washed with PBS and stained/fixed with methylene blue solution [2% w/v methylene blue (Sigma-Aldrich) in 70% Ethanol]. Fixed and stained colonies were then counted manually using a Stuart colony counter (Sigma-Aldrich).

2.7.4 γH2AX foci counting in 1BR cells after hTDP2 inhibitor and etoposide treatment

1BR human fibroblast cells were maintained as a monolayer culture in 75cm^2 flasks (T75, Corning) in modified Eagle's medium (α-MEM) media (Gibco) supplemented with 15% v/v Foetal Calf Serum (FCS), 1% v/v L-Glutamine (Gibco) and 1% v/v Penicillin-Streptomycin (10,000 U/mL) (Gibco). Cell culture was maintained in a NuAire Microbiological CO_2 incubator at 37°C. After reaching 90% confluence, cells were washed with PBS then trypsinised (as per EM9 cells, Section 2.7.2) to detach them from the flask and then counted using a haemocytometer.

25-30,000 cells/ml/plate were seeded in 15% v/v FCS containing media onto Nunc 4-well dishes (Thermo Fisher Scientific) containing round glass coverslips (neuVitro) to which the cells adhere. Cells were incubated at 37°C until they reached confluency (typically after 2 days), the media then was aspirated, and the coverslips washed with PBS. 1ml media containing 0.1% v/v FCS was added and the cells were incubated for an additional 5 days to induce cell cycle arrest.

Cell-cycle arrested cells were either pre-treated with 5uM hTDP2 inhibitor in 0.5% v/v DMSO or with 0.5% v/v DMSO as a vehicle control for 1 hour. After pre-incubation, 20μM etoposide (Sigma-Aldrich) in 0.5% v/v DMSO (final DMSO content was 1% v/v) was added to the cells. As a control, a set of cells was not subjected to VP16 treatment. After incubations for a further hour, the media was aspirated, and the coverslips washed with PBS; this step was set as the 'zero time' point. Control cells and one set of the treated cells were immediately fixed in 0.5ml 4% PFA (Paraformaldehyde)-PBS for 5 min at room temperature, and then washed with 2 successive applications of 0.5ml PBS. Fixed cells were stored at 4°C overnight in PBS. After washing the coverslips, fresh media was added back to the wells (pre-treated condition) and then incubated at 37°C. Cells were fixed at 30, 90 and 180 min time-points after VP16 treatment.

γH2AX foci counting: immuno-labelling protocol for 1BR cells

0.5ml PBS containing 0.2% v/v Triton-X100 was added to the wells containing the coverslips and then incubated for 5 min at room temperature. 0.5ml 5mg/ml BSA–PBS was then added and incubated for 1 hour at room temperature. 200μl 1mg/ml BSA–PBS buffer supplemented with 1:1000 dilution of mouse monoclonal anti-γH2AX antibody

(#2577, Cell Signalling Technology) and 1:1000 dilution of rabbit polyclonal anti-CENPF antibody (#ab5, Abcam) was added to the coverslips, and incubated for 1 hour at room temperature, followed by 3 successive applications of 0.5ml PBS containing 0.1% v/v Tween-20 and 0.2% v/v SDS (PBS-Tween-SDS) at room temperature. After that a 1:1000 dilution of Alexa Fluor 488-conjugated goat anti-mouse antibody (#A-11001, Thermo Fisher Scientific) and 1:1000 dilution of Alexa Fluor 555-conjugated goat anti-rabbit antibody (#A-21428, Thermo Fisher Scientific) in 1mg/ml BSA–PBS was added to the coverslips, and incubated for 1 hour at room temperature in the dark. Following this step, a 1:10,000 dilution of DAPI (Thermo Fisher Scientific) in PBS was added to the coverslips, which was then incubated for 5 min, followed by a final washing step with in PBS-Tween-SDS. The seeded cover slips were then mounted onto glass slides using VECTASHIELD mounting media (Vector Laboratories), and sealed with nail varnish. For foci counting a NIKON E400 microscope was used at 100x magnification. 40 cells arrested in G1 phase were chosen from each experimental set up to count γH2AX foci manually.

CHAPTER THREE

Expression, purification and crystallisation of human TDP2

3.1 Introduction

In order to pursue a structure-based drug design approach, one needs to first obtain reproducible high-resolution structural data for the protein in question. This approach typically requires a significant amount of protein at high purity that will be used to generate crystals, which will hopefully diffract to high resolution.

It was therefore necessary to first optimise the expression and purification of human TDP2. TDP2 is composed of two domains connected by a twenty amino acid interdomain loop reference. This loop represents a region of high flexibility that could potentially cause problems during crystallogenesis, therefore expression constructs containing either the full-length protein or catalytic domain (in isolation) were generated (Figure 3.1).

Expression trials of synthetic, codon-optimised full-length and catalytic domain hTDP2 constructs were undertaken in a bacterial expression system.

3.2 Results

A synthetic and codon-optimised gene encoding full length human TDP2 for expression in *Escherichia coli* (*E.coli*) was purchased from GenScript (Piscataway NJ, USA). The protein coding sequence was flanked by *Nde*I and *Eco*RI restriction sites to facilitate downstream sub-cloning into expression vectors.

3.2.1 hTDP2 expression trials in *E. coli*

Protein expression in bacterial expression systems has distinct advantages; it is fast and relatively cheap compared to other methods, but also has some associated problems: proteins of a mammalian origin will lack certain post-translational modifications, whilst multi-domain proteins may not fold properly. Protein expression trials usually begin in bacteria, but if sufficient recombinant protein is not produced, another expression host is often worth considering. For expression in *E.coli*, two expression plasmids were used: pTWO-E and pTHREE-E. pTWO-E is an in-house vector modified from pET-17b encoding an N-terminal human rhinovirus-3C cleavable (HRV-3C) His_6 affinity tag. pTHREE-E is an in-house vector modified from pGEX-6P-1 encoding an N-terminal HRV-3C cleavable glutathione-S-transferase (GST) affinity tag. Both plasmids were a kind gift of Dr A.W. Oliver, University of Sussex. DNA encoding both the full-length and catalytic domains of human TDP2 were sub-cloned into each expression plasmid.

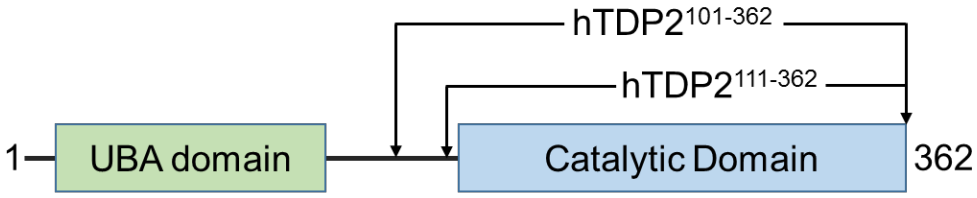

Figure 3.1. The domain structure of human TDP2. TDP2 contains two domains, an N-terminal ubiquitin-associated (UBA) domain and a C-terminal catalytic domain connected together by an approximately 20 amino acid interdomain loop. The amino acid boundaries of the two catalytic domain expression constructs are also indicated.

3.2.1.1 Expression trials of full length hTDP2

The *TDP2* open reading frame (ORF) was excised from the hTDP2-pUC57 plasmid by a *Nde*I / *Eco*RI restriction enzyme double digest and ligated into linearised, phosphatase-treated pTWO-E and pTHREE-E plasmids. Positive clones were verified by restriction enzyme double digest, followed by DNA sequencing. This sub-cloning step generated the expression vectors pTWO-E/hTDP2-FL and pTHREE-E/hTDP2-FL (Appendix A.1).

The pTWO-E/hTDP2-FL plasmid was introduced into *E.coli* strains BL21(DE3) and Rosetta(DE3) pLysS by transformation (see Materials and Methods 2.1.1). Rosetta(DE3) pLysS cells provide tighter control of protein expression as they also express T7 lysozyme, which reduces the basal level of expression from the T7 promoter found within the pTWO-E plasmid (Zhang and Studier 1997). Small-scale expression trials (5ml) were conducted by growing cells at 37°C or 20°C for 3 hours or overnight, respectively, with protein expression induced by the addition of 0.4mM IPTG, once the OD_{600} of the culture had reached 1.0. Protein extracts were prepared as described in Materials and Methods (2.2.6.1), and were analysed by SDS-PAGE using Novex 4-12% Tris-Glycine Protein Gels (Life Technologies, Paisley, UK) and staining of proteins with InstantBlue (Expedeon, Cambridge, UK). The expected molecular mass of polyhistidine-tagged full-length hTDP2 is 42 kDa.

The observed level of recombinant hTDP2 expression was much higher when cells were grown at 37°C than at 20°C, in both BL21(DE3) and Rosetta(DE3) pLysS cells. However, the recombinant protein was largely present in the insoluble fraction (Figure 3.2).

In order to confirm the presence of any hTDP2 in the soluble fraction, a larger 50ml culture was grown in L-broth, using BL21(DE3) cells, and induced for 3 hours at 37°C. Recombinant histidine-tagged hTDP2 was purified from the clarified supernatant using gravity flow column and immobilised metal affinity chromatography (IMAC) using TALON resin (Clontech, St-Germain-en-Laye, France). Whilst this approach resulted in the purification of some soluble hTDP2, the presence of a protein that migrated just below the 62kDa molecular weight marker suggested the presence of the bacterial chaperonin protein GroEL, which is highly indicative of misfolded protein (Figure 3.4A).

In parallel with the experiments discussed above, expression trials with a different affinity tag were undertaken, in an attempt to improve solubility. A glutathione S-transferase (GST) tag can help increase the solubility of a target protein and is readily captured from

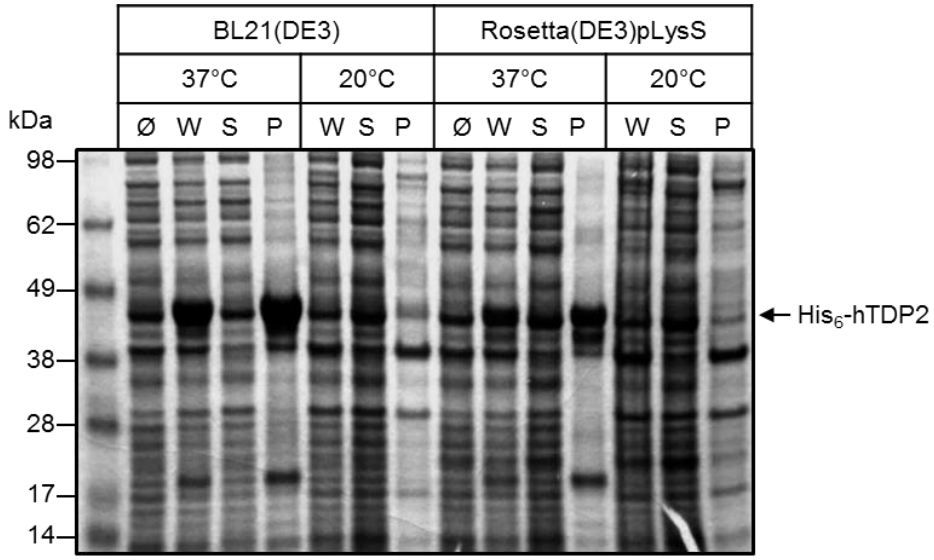

Figure 3.2. Small-scale expression trials of His-tagged full-length hTDP2. The indicated expression strain of *E. coli* was transformed with pTWO-E/hTDP2-FL plasmid. Expression trials were conducted in L-broth, at temperature of 20 or 37°C as indicated. Samples were run on Novex 4-12% Tris-Glycine Protein Gels and then stained with InstantBlue Protein Stain. KEY: (Ø): pre-induction sample, (W): whole cell lysate, (S): supernatant, and (P): pellet. The molecular weight (MW) of the full-length His_6-hTDP2 is 42kDa. An arrow indicates the expected size of the recombinant protein.

cell lysates by affinity chromatography. The pTHREE-E/hTDP2-FL plasmid was transformed into BL21(DE3) competent cells. As the pTHREE-E plasmid contains the *tac* promoter (a hybrid of the *trp* and the *lac* promoters), expression of the target protein is driven by the host's RNA polymerase, and not by T7 RNA polymerase – which means that there is no advantage in using pLysS strains of *E.coli* for tight control of protein expression. Small-scale expression trials (5ml) in L-broth were set up as before. After disruption of cells, and analysis of the soluble and insoluble fractions by SDS-PAGE, it became evident that most of the expressed protein was insoluble. However, given that there was a small amount of protein in the supernatant (Figure 3.3), this expression system was explored further.

The supernatant was loaded onto Glutathione Sepharose (GS) resin, unbound material removed by successive washes with lysis buffer, and the bound material recovered by boiling in SDS-PAGE loading buffer. SDS-PAGE revealed two bands of ~62kDa, again suggesting association of the recombinant TDP2 protein (~66kDa) with GroEL (~60kDa) (Figure 3.4B). To examine this further, I repeated the expression experiment as above, but using a larger 50ml culture. In addition, GST-hTDP2 was retained on the resin and subjected to overnight digestion by HRV-3C to remove the GST tag. The next day, fractions corresponding to the overnight digest, wash and bead fractions were analysed by SDS-PAGE. From this analysis, it was clearly obvious that hTDP2 and GroEL were still associated. The presence of co-purifying GroEL for both the His- and GST-tagged versions of full-length hTDP2 was verified by Western blot (Figure 3.4 C).

3.2.1.2 Expression trials of the catalytic domain constructs hTDP2-CAT$^{101-362}$ and hTDP2-CAT$^{111-362}$

In parallel with the experiments described in the previous section, attempts to clone and express just the catalytic domain of hTDP2 were undertaken (Materials and Methods 2.2.6.2) (Appendix A.2). To increase the likelihood of obtaining stable and soluble protein, two ORFs were generated by PCR, which differed in the number of N-terminal amino acids included, with hTDP2-CAT$^{101-362}$ encoding 10 extra amino-terminal amino acids than hTDP2-CAT$^{111-362}$. Both ORFs were sub-cloned into the expression plasmids pTWO-E and pTHREE-E (Table 3.1). pTWO-E/hTDP2-CAT$^{101-362}$ and pTWO-E/hTDP2-CAT$^{111-362}$ were transformed into BL21(DE3) and Rosetta(DE3) pLysS *E. coli* strains and expression tested in small scale experiments at different temperatures, induction times and growth media (Figure 3.5A and 3.5B, and Table 3.2). These set of experiments suggested

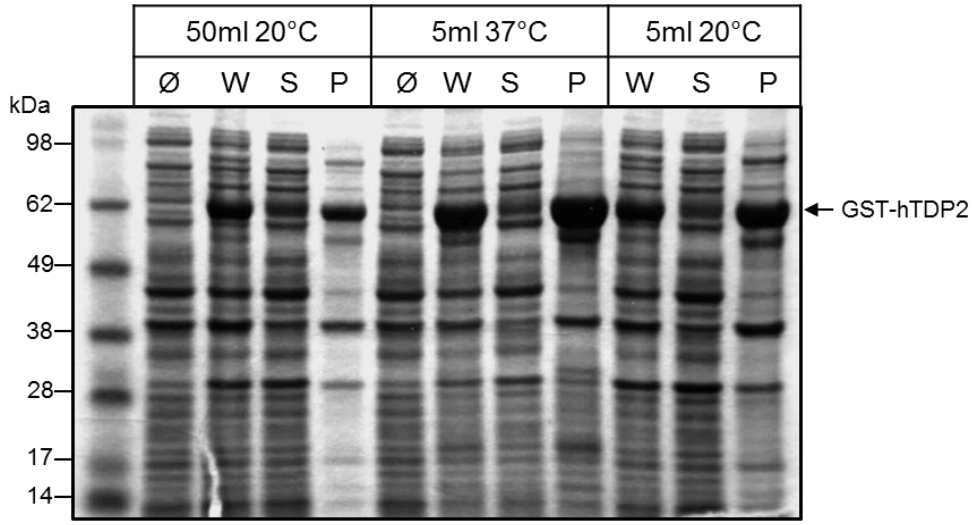

Figure 3.3. Small-scale expression trials of the GST-tagged full-length hTDP2. Expression of full-length hTDP2 was induced from pTHREE-E/hTDP2, encoding GST-tagged hTDP2. Expression trials were conducted in L-broth, at 20°C or 37°C as indicated. Samples were run in Novex® 4-12% Tris-Glycine Protein Gels (Life Technologies). Abbreviations: (Ø): pre-induction sample, (W): whole cell lysate, (S): supernatant, and (P): pellet. The MW of the full-length GST-hTDP2 is 66kDa. An arrow indicates the expected migration position of the recombinant protein.

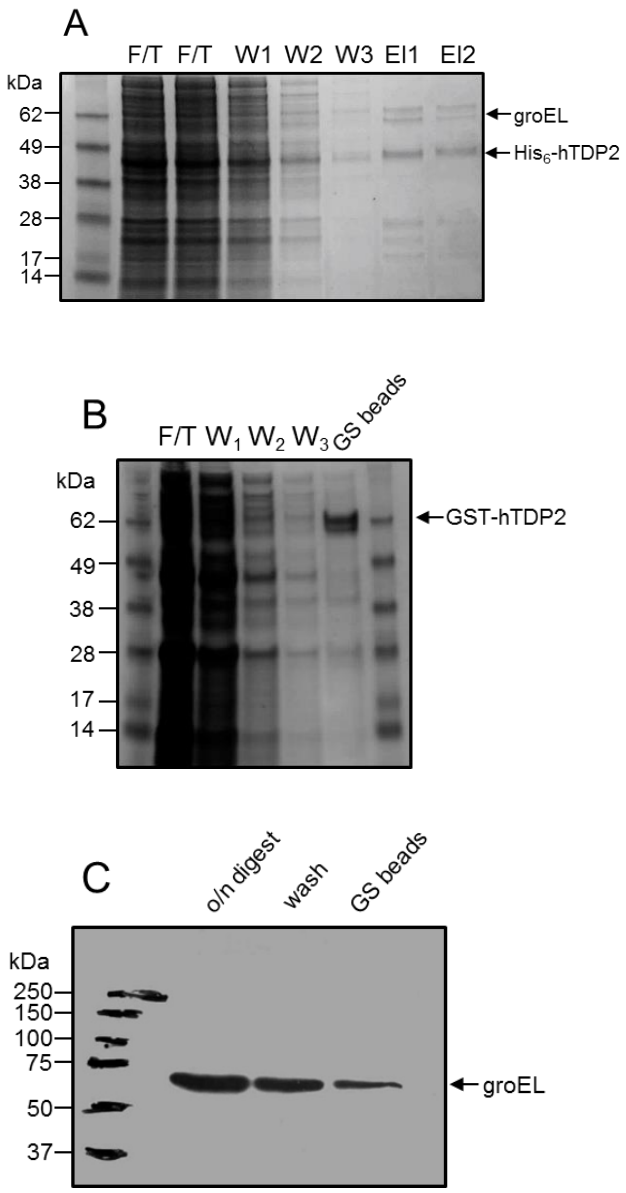

Figure 3.4. Full length TDP2 expressed in *E. coli* co-purifies with GroEL. **A.** His_6-tagged hTDP2-FL co-elutes with a ~60 kDa band. **B.** GST-hTDP2-FL also coelutes with a ~60 kDa band (both proteins run very close together on SDS-PAGE gels). **C.** A Western Blot, with an anti-GroEL primary antibody confirming the identity of the ~60 kDa band. KEY: (F/T): flow-through, (W): wash, and (El): elution.

Expression construct	Expressed protein
pTWO-E/hTDP2-CAT$^{101-362}$	His$_6$-hTDP2-CAT$^{101-362}$
pTWO-E/hTDP2-CAT$^{111-362}$	His$_6$-hTDP2-CAT$^{111-362}$
pTHREE-E/hTDP2-CAT$^{101-362}$	GST-hTDP2-CAT$^{101-362}$
pTHREE-E/hTDP2-CAT$^{111-362}$	GST-hTDP2-CAT$^{111-362}$

Table 3.1. Nomenclature of expression constructs and proteins. The pTWO-E constructs express an N-terminally His-tagged protein. The pTHREE-E constructs express an N-terminally GST-tagged protein. Both affinity tags can be removed by incubation with human rhinovirus-3C protease (HRV-3C).

that, of those conditions tested, overnight induction at 20°C in Turbo-broth (Turbo) media was best, in terms of the level of recombinant protein expression observed by SDS-PAGE. However, the solubility of the recombinant protein was still very poor, with most of the hTDP2 present as insoluble material in inclusion bodies. Similarly, pTHREE-E/hTDP2-CAT$^{101-362}$ and pTHREE-E/hTDP2-CAT$^{111-362}$ were transformed into BL21(DE3) *E. coli* strains and the expression tested in small scale experiments (Figure 3.5C). hTDP2 expression was the highest in Turbo-broth, for overnight induction at 20°C (Table 3.2), however the recombinant protein was again insoluble and accumulated in inclusion bodies.

3.2.2 Purification and crystallisation of hTDP2-CAT^{C161S}

During this period of troubleshooting, an expression construct encoding a mutated version of the catalytic domain (residues Met113 to Ile362) was kindly provided by Prof. Aidan Doherty (University of Sussex). The C161S mutation in this plasmid, denoted pET-28a/hTDP2-CAT^{C161S}, corresponds to C273S in the full-length human protein and had previously been shown by Prof. Doherty's group to improve the level of protein solubility, by removing a potentially reactive cysteine from the protein surface. This expression construct encodes an N-terminal His_6-tag and a thrombin cleavage site (Figure 3.6).

3.2.2.1 Expression and purification of hTDP2-CAT^{C161S}

For initial small-scale expression trials (50ml), expression was tested at different temperatures (37°C and 20°C) and in different media (L-broth and Turbo broth). The supernatant was loaded onto batch gravity flow column for IMAC capture with TALON resin, and the eluted protein samples were analysed by SDS-PAGE (Figure 3.7). Based on this experiment, an optimal expression condition, comprising 3.5 hours induction at 30°C, with cells grown in Turbo broth was determined. The detailed purification protocol is described in Materials and Methods (2.2.6.3). In brief, hTDP2 protein was purified in three steps, first recovering TDP2 from the clarified supernatant by IMAC (TALON resin) followed by salt fractionation on a Heparin Sepharose 6 Fast Flow column (GE Healthcare) (Figure 3.8 A), followed by size exclusion chromatography (Figure 3.8 B). Peak fractions containing hTDP2 were pooled, concentrated to 7.3mg/ml, and stored at -80°C.

3.2.2.2 Crystallisation of hTDP2-CAT^{C161S}

Initial screens, in 96-well, sitting-drop vapour-diffusion format (MRC2 plates, Hampton Research, Aliso Viejo, CA USA) were set up using commercial crystallisation screens purchased from Hampton Research; these included JCSG+, Index, PEG/Ion and Salt-RX

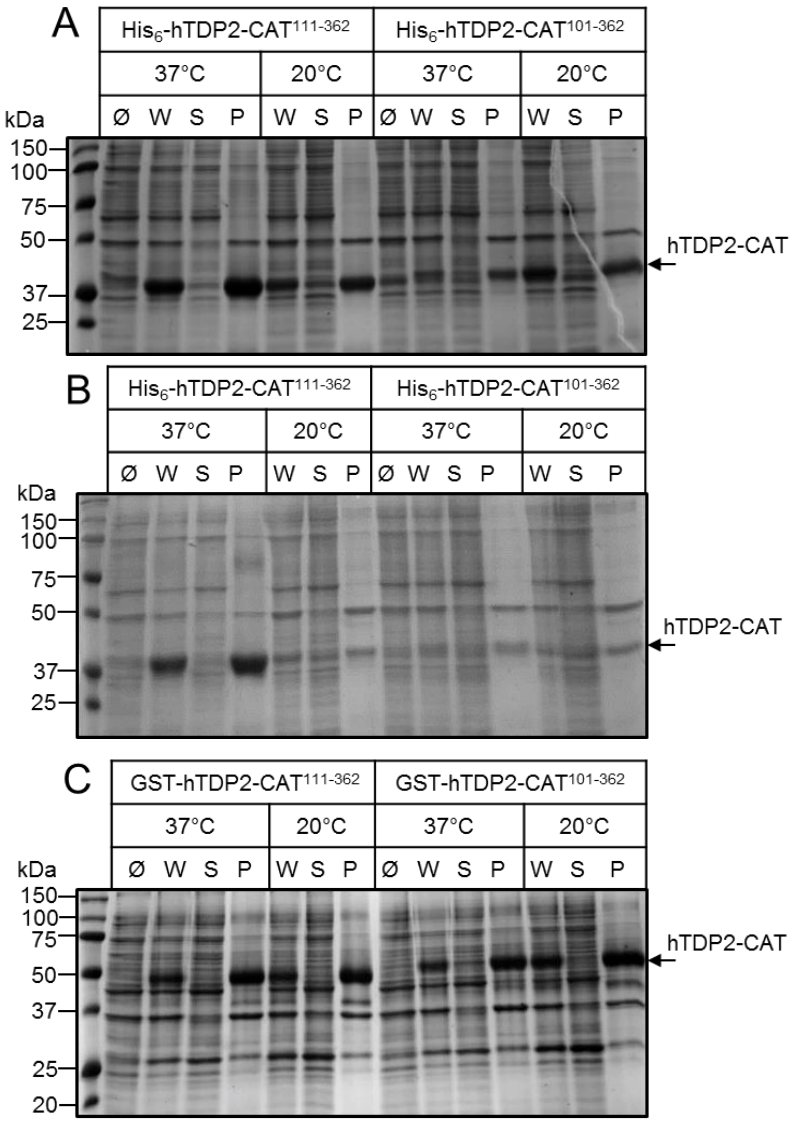

Figure 3.5. Small scale expression trials for pTWO-E and pTHREE-E / hTDP2-CAT$^{101-362}$ and hTDP2-CAT$^{111-362}$. Cell cultures were grown in L-broth, and grown at 20˚C overnight, after induction of protein expression. **A. & B.** Expression trials for pTWO-E constructs in *E. coli* strains BL21, and Rosetta2(DE3) pLysS, respectively. **C.** Expression trials for pTHREE-E constructs in *E. coli* strain BL21. The different experimental temperatures tested are as indicated at the top of each SDS-PAGE gel. An arrow in each case indicates the expected migration position of the recombinant protein. KEY: (Ø): pre-induction sample, (w): whole cell lysate, (s): supernatant, and (p): pellet.

Exp. construct	hTDP2-CAT$^{101-362}$									hTDP2-CAT$^{111-362}$								
vector	pTWO-E (His_6 tag)						pTHREE-E (GST tag)			pTWO-E						pTHREE-E		
E. coli strain	BL21(DE3)			Ros.(DE3)pLysS			BL21(DE3)			BL21(DE3)			Ros.(DE3)pLysS			BL21(DE3)		
Exp. T	37°C	20°C		37°C	20°C		37°C	20°C		37°C	20°C		37°C	20°C		37°C	20°C	
Media	LB	LB	TB	LB	LB	TB	LB	LB	TB	LB	LB	TB	LB	LB	TB	LB	LB	TB
Level of exp.	*	**	***	Ø	Ø	**	*	**	***	***	**	*	***	*	***	*	*	***
S/N or P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P

Table 3.2. An overview of all small-scale expression trials conducted for hTDP2-CAT$^{101-362}$ and hTDP2-CAT$^{111-362}$. Both catalytic domain coding DNA sequences were cloned into the expression plasmids pTWO-E and pTHREE-E. The level of expression is represented by asterisks, where * = low level expression; ** = medium level expression, and *** = high level expression, and Ø = no expression. KEY: Ros.(DE3)pLysS: Rosetta(DE3)pLysS, (LB): L-broth, (TB): Turbo-broth, (S/N): supernatant and (P): pellet.

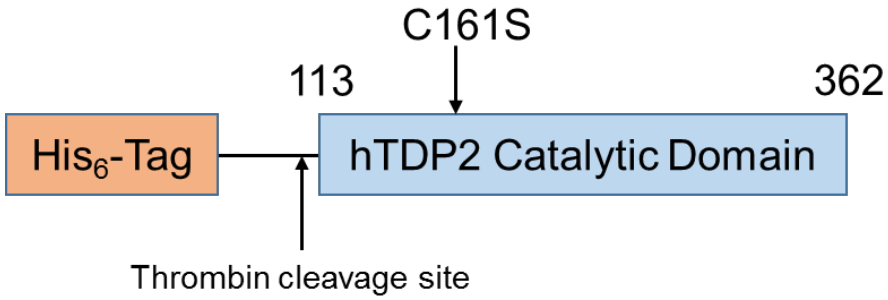

Figure 3.6. Schematic representation of pET-28a/hTDP2-CAT^{C161S}. The catalytic domain of hTDP2 (residues spanning from Met113 to Ile362) was cloned into the pET-28a expression vector containing an N-terminal His_6-tag and a thrombin cleavage site. The lone cysteine at position 161 was mutated to serine in order to prevent putative intermolecular disulphide bond formation.

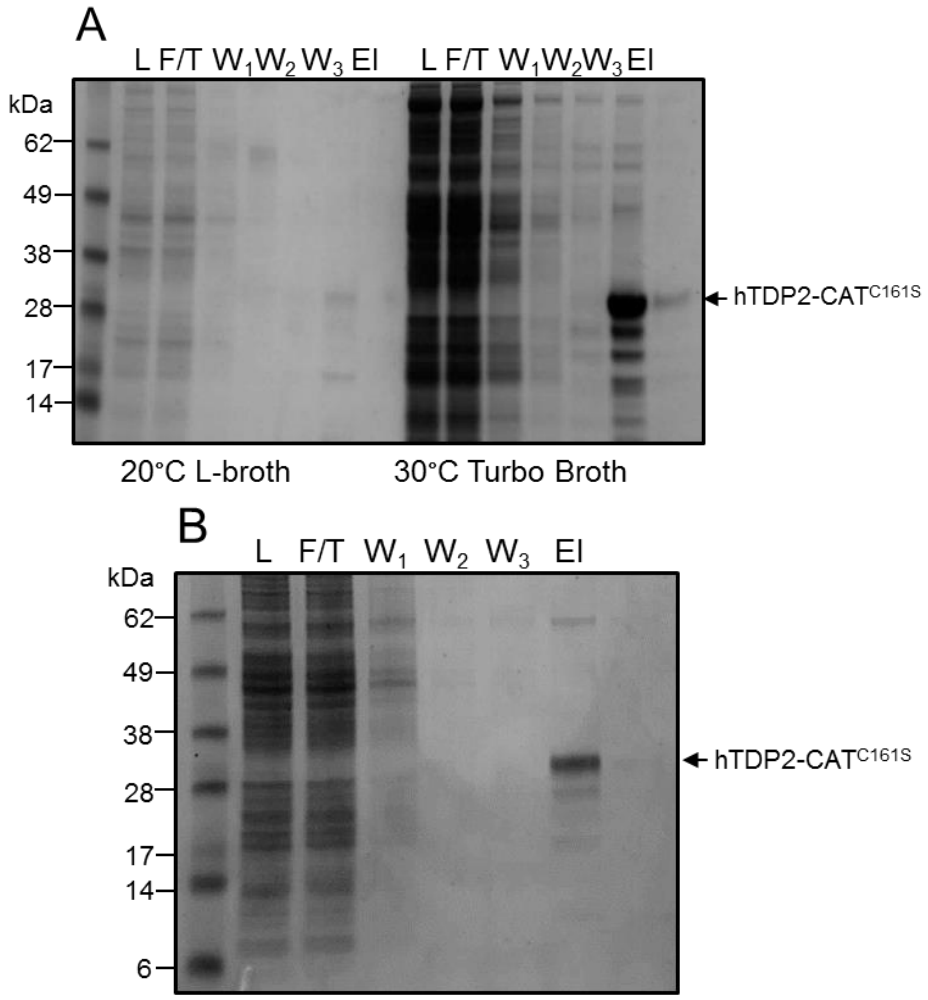

Figure 3.7. Small scale expression trials and IMAC pull-downs of hTDP2-CAT^{C161S} expressed in *E. coli* strain B834(DE3). **A.** Initial experiments were carried with overnight growth, after induction of protein expression, at different temperatures and in different growth media. **B.** Optimised expression condition, with IMAC pull-down: 3.5-hour induction at 30°C in Turbo broth. KEY: (L): lysate, (F/T): flow-through, (W): wash, and (El): elution. The MW of poly-Histidine tagged hTDP2-CAT$^{161-406}$ is 28kDa. An arrow indicates the expected migration position of the recombinant protein.

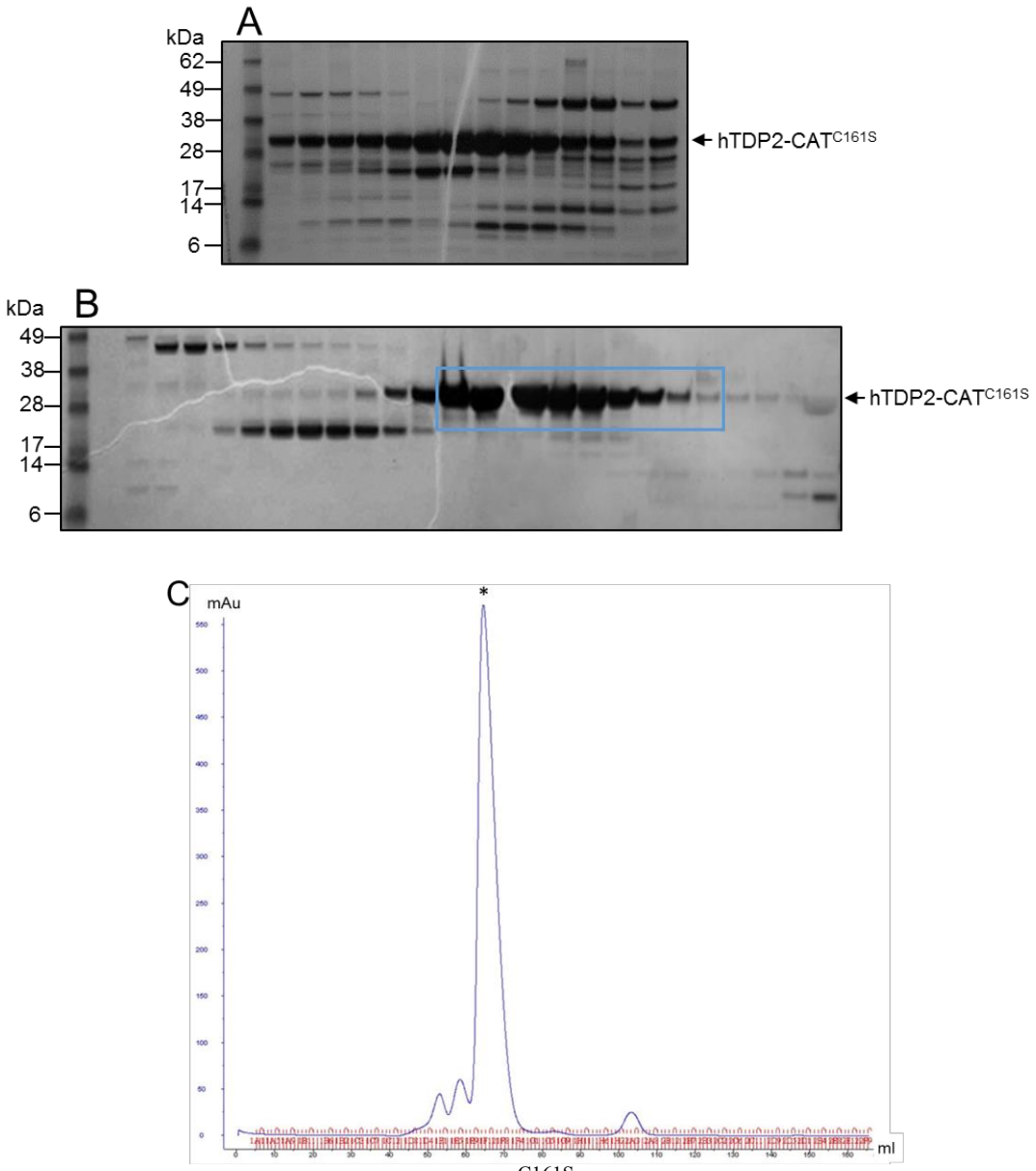

Figure 3.8. Purification of hTDP2-CAT^{C161S}. **A.** Following initial capture by IMAC (TALON resin), the recombinant protein was loaded on to a Heparin Sepharose 6FF column. Bound material was eluted with the application of a linear salt gradient, from low to high NaCl concentration. A representative SDS-PAGE gel for fractions collected during the elution gradient is shown. **B.** Fractions containing hTDP2-CAT^{C161S} (blue box) were identified, pooled and concentrated, then loaded onto a HiLoad 16/600 Superdex 75pg size exclusion column. A representative SDS-PAGE gel for fractions collected during purification is shown. The expected migration position of the recombinant protein is marked by an arrow in each case. **C.** A representative chromatograph for the size exclusion chromatography purification step.

screens. Screens were checked regularly for the presence of crystal growth, and after a period of five days, small needle-shaped crystals were seen to emerge in several different conditions (G5, G6, A10 and A11 from Salt RX and condition B4 from Index screens) (Figure 3.9 and Table 3.3).

Crystals obtained in the A10 SaltRx (Hampton Research), looked to be promising, so were mounted in a cryoloop, swiped through the mother liquor (2.2M NaCl, 0.1M bis-tris propane pH7.0) containing various cryoprotectants (20% ethylene glycol, or 30% glucose, or 40% sucrose) to protect the crystals, and then shot in-house on a MicroMax-007HF rotating anode source (Rigaku, Sevenoaks, UK). Diffraction was observed, but only to low (8-9Å) resolution. It was therefore necessary to try to optimise these initial crystallisation conditions.

A grid-optimisation method, in a hanging-drop vapour-diffusion format using 24-well plates [VDX Plate with sealant (Hampton Research)] was used. Crystallisation buffers were kept at a constant concentration of 0.1M, but pH was varied by 0.5 unit increments between pH7.0 and pH8.5. As TDP2 is known to require magnesium for its catalytic activity, it was therefore thought that having this ion in the crystallisation buffer (as observed in G5 and G6 conditions of SaltRx and B4 of Index screens) might help to stabilise the protein and to promote crystallisation. Different concentrations of both NaCl and $MgSO_4$ were screened in 200mM increments (1.8-2.8M for NaCl and 1-2M for $MgSO_4$).

From the grid-optimisations, it became apparent that using 0.1M bis-tris propane as a buffer at a pH of 7.0 generated larger crystals in the presence of high concentrations of NaCl (1.8M), albeit that these crystals were elongated in only one dimension.

In a further attempt to improve crystal quality, several different magnesium salts (sulphate, acetate, chloride and formate) were screened in the presence of 0.1M bis-tris propane pH7.0 across different NaCl concentrations. The presence of 5mM magnesium acetate promoted crystal growth when the concentration of NaCl decreased to 0.45M.

The condition 0.1M bis-tris propane pH7.0, 0.45M NaCl, 5mM Mg acetate provided reasonably large needle-shaped crystals. Again, these crystals were cryo-protected, and then shot in-house. The resulting diffraction was highly anisotropic, ranging from 2.5Å to

Conditions from SaltRx screen	
A10	2.2M NaCl, 0.1M bis-tris propane pH7.0
A11	2.2M NaCl, 0.1M Tris pH8.5
G5	1.8M $MgSO_4$, 0.1M bis-tris propane pH7.0
G6	1.8M $MgSO_4$, 0.1M Tris pH8.5
Condition from Index screen	
B4	0.3M Mg formate, 0.1M Tris pH8.5

Table 3.3. Initial crystallisation 'hits' with commercial screens from Hampton Research. Sitting-drop vapour-diffusion format (MRC2 plates, Hampton Research). Screens were set up using the Phoenix liquid handling system.

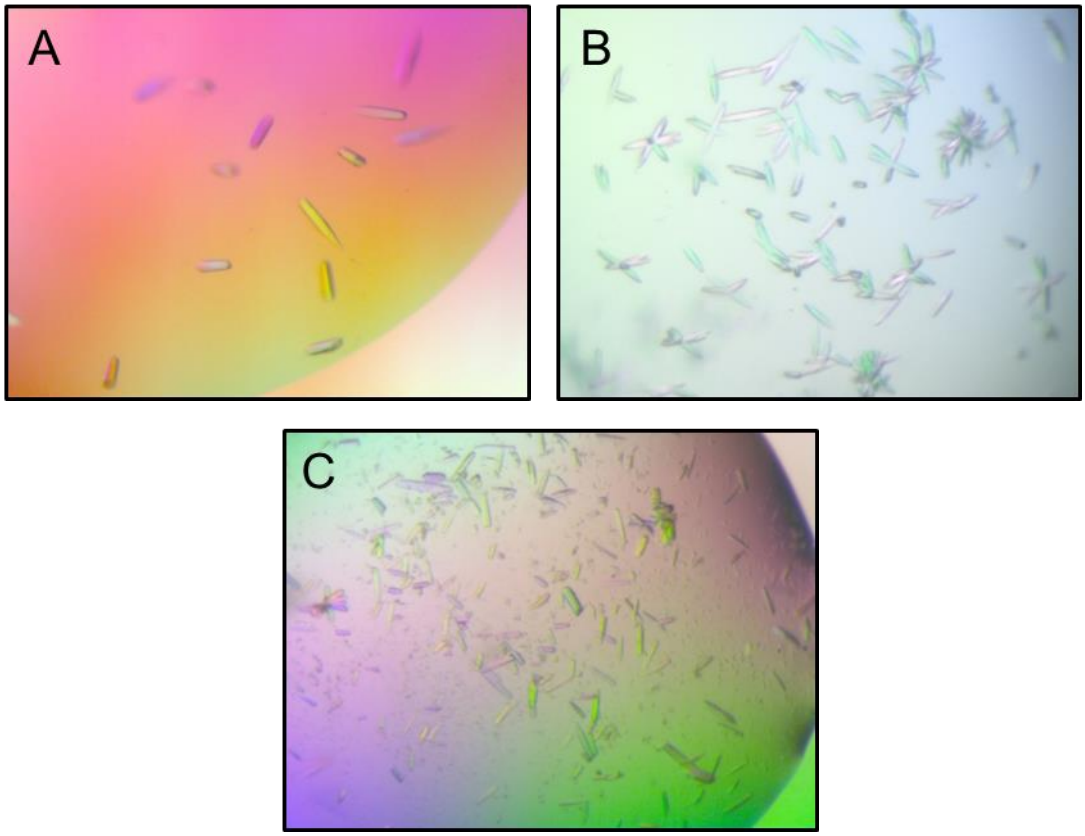

Figure 3.9. Images of 'hits' from initial crystallisation screens with hTDP2-CAT^{C161S}. **A.** SaltRx condition A10: 2.2M NaCl, 0.1M bis-tris propane pH7.0. **B.** SaltRx condition G5: 1.8M $MgSO_4$, 0.1M bis-tris propane pH7.0. **C.** Index condition B4**:** 0.1M Tris pH8.5, 0.3M Mg formate.

3.8Å, depending on crystal orientation. The likely source of anisotropy in this case, was due to the morphology of the TDP2 crystals, which were long and thin.

In order to improve the observed anisotropy, an additive screen from Hampton research was employed (Cat no: HR2-138). The additive screen is a library of small molecules and can alter sample-sample and sample-solvent interactions and can improve crystal formation. From this screen, it was found that the additive trimethylamine *N*-oxide (TMAO) appeared to improve crystal morphology. Therefore, an extended optimisation screen, which varied the concentration of TMAO in 0.5% v/v increments (ranging from 0 to 2.5%) was undertaken.

Crystals grown in 0.1M bis-tris propane pH7.0, 0.45M NaCl, 5mM Mg acetate + 1.5 % v/v TMAO were shorter in length and appeared much thicker (Figure 3.10) (Materials and Methods 2.4.1.1). These crystals were harvested, then plunge-frozen in liquid N_2 in the presence of 20% v/v ethylene glycol, 40% w/v sucrose or 30% v/v glycerol to test the effectiveness of different cryoprotectants. An isotropic diffraction dataset to 3.1Å resolution was collected on beamline I04 at the Diamond Light Source (DLS, Didcot, UK) (Table 3.4).

3.2.2.3 Expression, purification and crystallisation of Selenomethionine labelled hTDP2-CAT^{C161S}

To obtain phase information, we decided to use the single-wavelength anomalous dispersion method (SAD) with Selenomethionine-derivatised (SeMet) recombinant protein. As the expression construct contained 10 methionine residues, we surmised that sufficient phasing power would be generated with this methodology. At this time point, there were no available structures for TDP2 deposited in the Protein Data Bank (PDB), therefore molecular replacement could not be used to solve the phase problem.

The pET-28a/hTDP2-CAT^{C161S} expression construct was transformed into an auxotrophic strain of *E.coli,* B834(DE3) – which is dependant on an external source of methionine for growth. Transformed *E.coli* cells were first selected on minimal media-agar plates containing methionine, supplemented with 50µg/ml kanamycin. Small-scale 'starter cultures' were grown in LB media. Large-scale expression cultures were then grown in Selenomethionine containing media using SelenoMet media from Molecular Dimensions (Materials and Methods 2.2.6.4.1). Expression and purification protocols for the

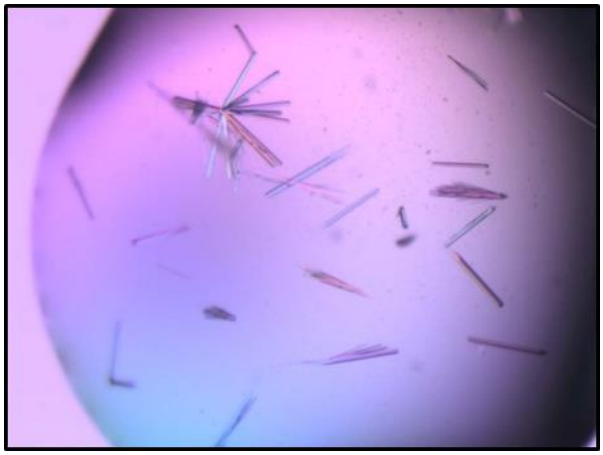

Figure 3.10. Image of optimised hTDP2-CAT^{C161S} crystals. Hanging drop vapour-diffusion format. Crystallisation condition: 100mM bis-tris propane pH7, 0.45M NaCl, 0.05M $MgAc_2$, 1.5% v/v trimethylamine *N*-oxide. Crystals appeared after 5 days of incubation at 20ºC. Crystals were frozen in the presence of various cryoprotectants (20% ethylene glycol, or 30% glucose, or 40% sucrose) of which hTDP2 crystal with 40% sucrose diffracted to 3.1Å.

Space group	P3
Unit cell dimensions	91.5Å; 91.5Å; 118.0Å
Internal angles	90°; 90°; 120°

Table 3.4. Space group and cell dimensions of hTDP2-CAT^{C161S} crystals. Crystallisation condition: 100mM bis-tris propane pH7, 0.45M NaCl, 0.05M $MgAc_2$, 1.5% v/v trimethylamine *N*-oxide. Crystals appeared after 5 days of incubation at 20ºC. Crystals were snap-frozen in liquid N_2, after being cryo-protected in in the presence of crystallisation buffer supplemented with 40% v/v sucrose. Diffraction data were collected on beamline I04, at DLS.

Selenomethionine derivatised hTDP2-CAT^{C161S} protein are identical to those for the unlabelled protein and can be found in Materials and Methods (Section 2.2.6.4)

Crystallisation conditions for the derivatised protein were also similar: crystals were grown in a hanging drop vapour-diffusion format, at a temperature of 20°C. 1 µl of SeMet-labelled protein was combined with 1 µl of crystallisation buffer (100mM Bis-Tris propane pH7.0, 0.5M NaCl, 0.05M $MgAc_2$, 1.5% v/v trimethylamine *N*-oxide) (Materials and Methods 2.4.1.3). Small rod-shaped crystals appeared after about 5 days (Figure 3.11) and were snap-frozen in liquid N_2, after being cryo-protected in crystallisation buffer supplemented with 30% v/v glycerol.

Diffraction data were collected on beamline I04, at DLS. However, only a poor, low resolution dataset was collected which precluded phasing by the SAD method.

However, at this time point the structure of the catalytic domain of murine TDP2 (mTDP2-CAT) was published (Schellenberg et al. 2012). By using the atomic coordinates of the murine protein in molecular replacement, it became possible to solve the structure of the catalytic domain of human TDP2 (Figure 3.12).

3.3 Discussion

The aim of this set of experiments was to express and purify hTDP2 in large enough quantities to facilitate crystallography, and ultimately enable structure-based drug design efforts, in order to produce inhibitors of TDP2 enzymatic function.

Full length TDP2 is composed of two domains, connected by an approximately forty amino acid long linker region. Because this connecting loop was likely to have a high degree of flexibility, which might cause problems during crystallization, constructs expressing just the catalytic domain of hTDP2 were also created.

In expression trials with full-length hTDP2, IMAC pull-downs showed that some soluble protein was produced, but that this protein always co-purified with the bacterial chaperone GroEL, suggesting misfolding and aggregation. Similarly, in expression trials with the catalytic domain, all the expression conditions tested resulted in accumulation of TDP2 in the insoluble fraction. It is possible to purify proteins from the insoluble fraction / inclusion bodies in denaturing conditions. However, obtaining active and correctly folded protein from inclusion bodies can be an arduous task, although it does have some distinct

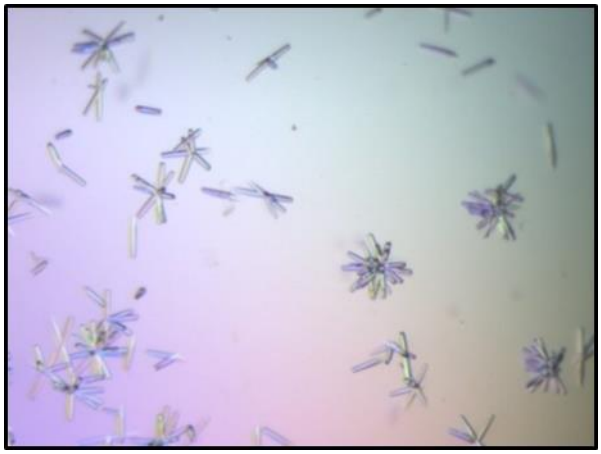

Figure 3.11. Image of Selenomethionine-labelled hTDP2-CAT^{C161S} crystals. Crystallisation condition: 100mM bis-tris propane pH7.0, 0.45M NaCl, 0.05M $MgAc_2$, 1.5% v/v trimethylamine *N*-oxide. Crystals were snap-frozen in liquid N_2, after being cryo-protected in the presence of crystallisation buffer supplemented with 30% v/v glycerol.

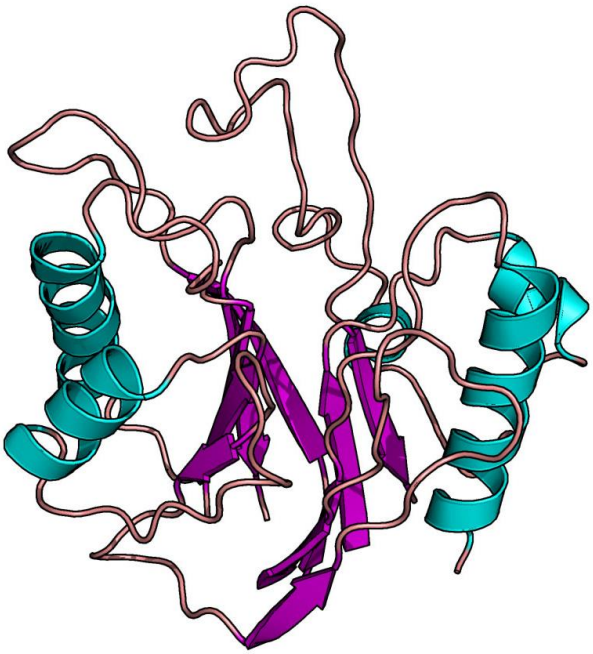

Figure 3.12. Carton representation of the catalytic domain of hTDP2. The structure was solved by molecular replacement using the catalytic domain of the murine TDP2 as a search model (PDB accession code: 4GYZ, Schellenberg et al. 2012).

advantages, as the protein is already in a semi-pure state, highly concentrated and generally protected from the action of bacterial proteases (Palmer and Wingfield 2004).

However, at this time point, an expression construct was kindly provided by Prof. Aiden Doherty. They had determined that a single point mutation of Cysteine 161 to Serine (C161S), found on the surface of the hTDP2 catalytic domain, enabled correct folding of the protein, which could now be readily purified from the soluble fraction. After initial expression trials in *E. coli*, it was possible to express a substantial amount of soluble protein which could be purified to high degree. As an alternative strategy, in overcoming the problems of reactive surface cysteines, one can express recombinant protein in a more reductive environment, i.e. employing strains deficient in thioredoxin reductase (Rosano and Ceccarelli 2014), or as in this case, single point mutations can be introduced in order to change the problematic surface residues.

Preliminary crystallisation trials resulted in crystals that only diffracted to low resolution. Therefore secondary optimisation screens were used to try and improve the crystals. Protein crystallisation depends on many factors (e.g. buffer and salt conditions, temperature), in the case of hTDP2, changing the pH of the buffer to 7.0 and including the magnesium co-factor Mg^{2+} resulted in a considerable improvement in the resolution of diffraction; however, this was highly anisotropic.

Several factors can contribute to anisotropy (Sawaya 2014). The crystal itself is a field of anisotropicity, movement of whole molecules as rigid bodies, and vibration along torsion angles can also be the source of anisotropy. For hTDP2, the major factor was simply the physical dimensions of crystals (elongated and thin crystals, 150µm x 15µm).

Additive screens were used to try to improve crystal quality. One additive, trimethylamine *N*-oxide, resulted in improved, isotropic diffraction of hTDP2 crystals to 3.1Å resolution. However, several crystals still had to been screened in order to obtain this dataset.

Once diffraction data was obtained, a method for solving the phase problem was required. We initially decided to use the single-wavelength anomalous dispersion method (SAD), using Selenomethionine derivatised protein. Whilst crystals of this protein were obtained, they did not diffract to a sufficient resolution to allow the structure to be solved. However, molecular replacement became a viable option when the structure of the catalytic domain of the murine TDP2 was published and made available (Schellenberg et al. 2012).

CHAPTER FOUR

Small molecule inhibitors of TDP2

4.1 Introduction

Selective inhibitors of hTDP2 were identified and described, as part of a collaboration between the Caldecott group and the Paterson Institute for Cancer Research (Raoof et al. 2013). These compounds were identified in a high-throughput screen using a chromogenic enzymatic assay (Thomson et al. 2013).

The report described a series of toxoflavin inhibitors, plus a single deazaflavin, which had a sub-micromolar inhibitory effect on the enzymatic action of hTDP2. However, further investigations by this team lead to the discovery of redox liability and *in vitro* drug metabolism for the toxoflavin series, making them unsuitable for progression. However, despite being a singleton hit, the deazaflavin class did not have the same undesirable properties as the toxoflavins. It was therefore selected for progression, and generated a series of inhibitors, which provided some limited structure-activity relationship (SAR) data.

Several TDP2 inhibitors were provided by the Cancer Research UK Drug Discovery Unit at the Paterson Institute for Cancer Research (University of Manchester, UK). These compounds were first tested in a range of additional assays before co-crystallisation attempts were made with recombinant hTDP2.

4.2 Results

4.2.1 Thermal denaturation assays to investigate compound-binding

A thermal denaturation assay, or ThermoFluor assay, is a useful tool to investigate both protein folding and stability. The protein in question is denatured by increasing temperature in the presence of an environmentally sensitive fluorescent molecule (in our case: SYPRO Orange, Sigma-Aldrich). During denaturation, hydrophobic amino acid residues normally found in the core of the protein are exposed to solvent. SYPRO Orange is able to bind to these exposed hydrophobic residues, leading to an increase of fluorescence, which can be measured. The temperature midpoint (Tm) of the folded-to-unfolded transition for the protein can also be monitored in the absence/presence of a compound that may bind to the protein in question. Binding of a compound (to the active site of the protein, if an enzyme) can affect the measured Tm point, typically resulting in a

positive shift to a higher temperature (although negative shifts, i.e. destabilisation of the protein are often seen). Thermal denaturation assays can therefore be used as a high throughput method to test if a particular compound interacts / binds to a protein (Lo et al. 2004, Senisterra et al. 2012).

Firstly, the suitability of hTDP2-CAT^{C161S} for this particular type of assay had to be determined; that is, confirming if it unfolded with a preferred sigmoidal shaped denaturation profile. Secondly, as compounds often have limited aqueous solubility and require organic solvent to remain in solution [3% v/v dimethyl sulfoxide (DMSO) is typically sufficient], the tolerance of the protein to increasing concentrations of DMSO was also tested. Thirdly, and finally, ten putative TDP2 inhibitor compounds were tested as a dilution series with the optimised assay. The assay is described in detail in Materials and Methods (section 2.6.1).

The suitability of the recombinant hTDP2-CAT^{C161S} protein was tested in two experimental buffers: HEPES pH7.0 and bis-tris propane pH7.5. Bis-tris propane was selected as it appeared in several crystallisation conditions that produced TDP2 crystals. From these initial experiments, a Tm of 47.6 °C was calculated for the protein in HEPES buffer. The Tm in bis-tris propane buffer was slightly higher, at 49.7°C (Figure 4.1). The HEPES buffer was selected for ongoing experiments, as it was felt that the lower Tm would be beneficial for screening, as compounds producing only moderate shifts in Tm might be overlooked in experiments with bis-tris propane; i.e. maximising the potential ΔTm.

To address the impact of DMSO on the assay, the stability of hTDP2 in a dilution series of DMSO was also examined. hTDP2 was found to be tolerant to concentrations of up to 3% v/v DMSO, as judged by this assay, as the observed Tm (47.6°C) did not change in the presence of DMSO (Figure 4.2).

Finally, deazaflavin inhibitors of hTDP2 were also tested in this assay format. Depending on the solubility of each compound, a dilution series ranging from ~ 50μM to ~ 300μM was used. From the 10 compounds supplied by the Paterson Group, two candidates produced a significant positive shift in Tm (Table 4.1). Compound C13984 had a ΔTm of +2.8°C at a concentration of 333μM, whereas C14005 had a ΔTm of +3.5°C at 300μM (Figure 4.3). Interestingly, one compound (C14313) clearly destabilised the protein, producing a ΔTm of -6.5°C.

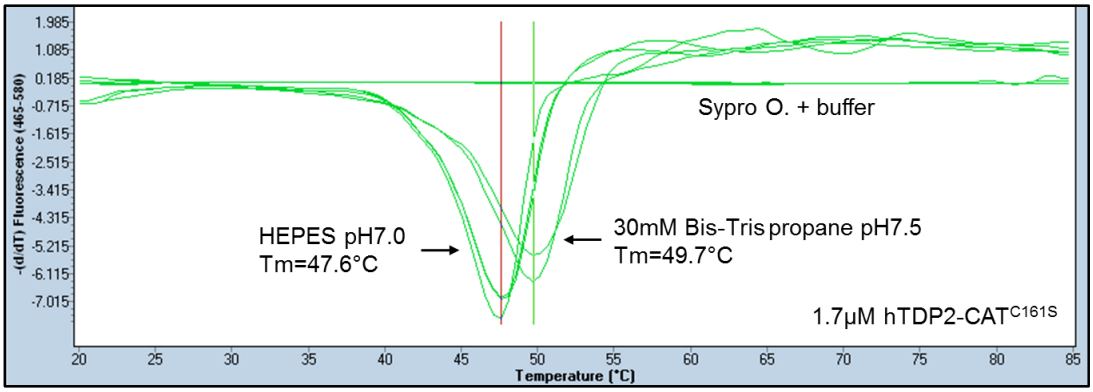

Figure 4.1. Thermal denaturation assay of hTDP2-CAT^{C161S} in different buffer systems. First order derivatives of the observed melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Reaction conditions: HEPES pH7.0 (or bis-tris propane pH7.5), 300mM NaCl, 0.5mM TCEP.

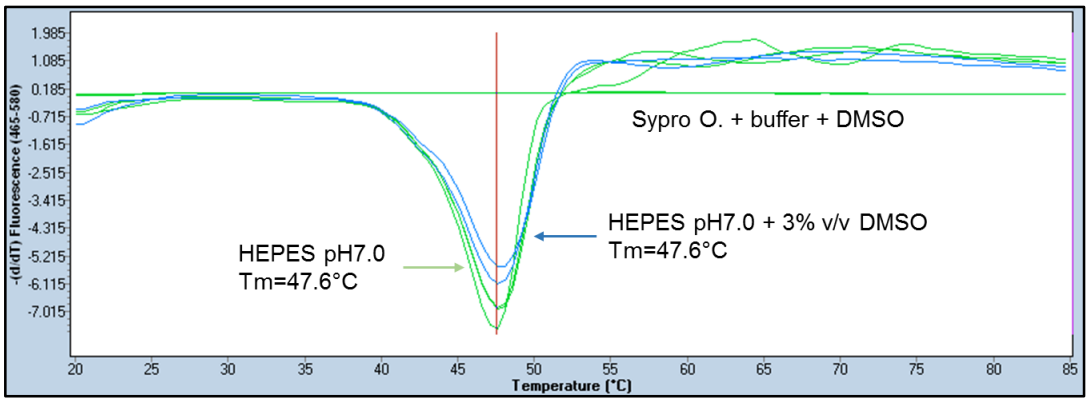

Figure 4.2. Thermal denaturation assay of hTDP2-CAT^{C161S} in various concentrations of DMSO. First order derivatives of the observed melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Reaction conditions: HEPES pH7.0, 300mM NaCl, 0.5mM TCEP and 0 or 3% v/v DMSO.

Compound ID	DMSO Stock Conc. (mM)	Thermal denaturation assay	Comments
C2062	25mM	Tested at 333 µM	Complete denaturation
C2216	25mM	Tested at 333 µM	Complete denaturation
C2303	25mM	55, 111, 167, 222, 333 µM	ΔTm: 0°C
C4231	Max conc.: 3,5mM	Tested at 105 µM	Reacts with Sypro/Denatures protein
C13834	Max conc.: 7,5mM	Tested at 250 µM	ΔTm: +0.7°C
C13984	25mM	55, 111, 167, 222, 333 µM	ΔTm: +2.3°C, good
C14005	Max conc.: 10mM	50, 100, 150, 200 300 µM	ΔTm: +3.5°C very good
C14313	25mM	Tested at 333 µM	ΔTm: -5.8°C
C14369	25mM	Tested at 333 µM	Reacts with Sypro O. / Denatures protein
C14370	25mM	Tested at 333 µM	Reacts with Sypro O. / Denatures protein

Table 4.1. Summary of results for thermal denaturation assays (hTDP2- CAT^{C161S} in the presence of different deazaflavin compounds). Compounds were provided by the Cancer Research UK Drug Discovery Unit, Patterson Institute for Cancer Research (University of Manchester). Stock solutions for each compound were prepared in 100% v/v DMSO due to their limited aqueous solubility. Thermal denaturation assays were performed at the indicated concentration of compound in the presence of 1.7µM hTDP2-CAT^{C161S}.

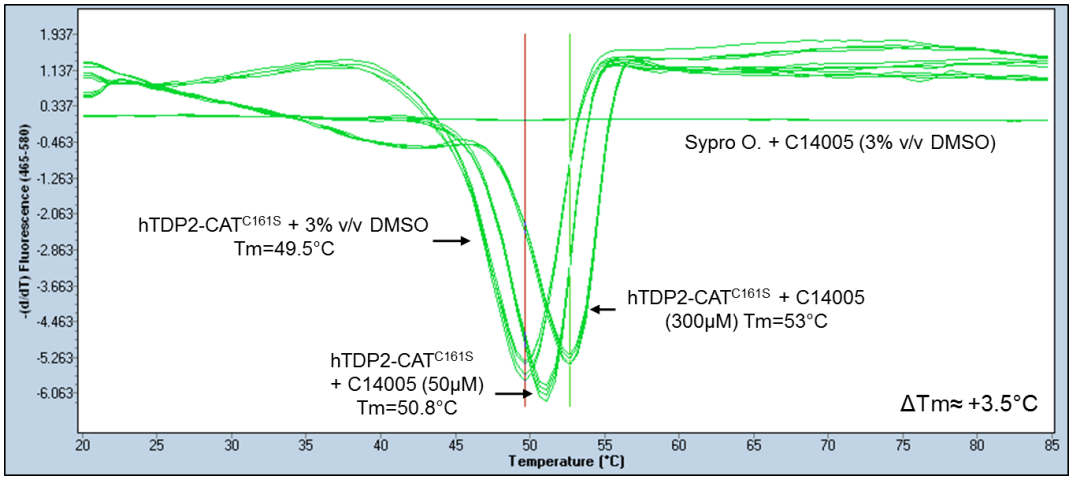

Figure 4.3. Thermal denaturation assay of hTDP2-CAT^{C161S} in the presence of compound C14005. First order derivatives of the observed melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Final compound concentration: 300µM. Reaction conditions: HEPES pH7.0, 300mM NaCl, 0.5mM TCEP and 3% v/v DMSO.

As C14005 – a phenylmethanesulfonamide derivative – produced the largest positive shift in Tm, it was selected for further biochemical experiments.

4.2.2 TDP2 enzymatic assay in the presence of compound C14005

The catalytic activity of recombinant hTDP2-CAT^{C161S} enzyme was measured in the absence and presence of C14005. The DNA substrates employed in this gel-based assay were [α-^{32}P]-labelled oligonucleotide duplexes containing a 5' overhang (4 nt) harbouring a 5'-phosphotyrosine moiety (Figure 4.4A). The experimental approach was to (a) test the enzyme activity of the recombinant enzyme, (b) test the impact of 3% v/v DMSO on activity, and (c) test activity in the presence of C14005. As mentioned in the previous section, the compounds require the presence of an organic solvent (in our case DMSO) in order to stay in solution. The essay is described in detail in Materials and Methods (section 2.3.2).

hTDP2-CAT^{C161S} at two different concentrations (6ng and 12ng/reaction) was tested with the labelled substrate withdrawing aliquots from the enzymatic reaction at 15, 30 and 60 minute time points. Under these experimental conditions, the reaction was completed within 15 min. The protein concentration was therefore reduced to 10nM and the time course was reduced to 15 min in order to observe comparable changes in activity (data not shown).

The impact of DMSO on the assay (Figure 4.4B) was also investigated. Experiments revealed that addition of 3% v/v DMSO had no observable effect on the activity of the enzyme; therefore, the effect of C14005 on the enzymatic activity of hTDP2-CAT^{C161S} was studied.

Preliminary experiments failed to reveal any inhibitory effect of C14005 on hTDP2 activity (data not shown). It was thought that some components of the reaction buffer might be interfering with compound binding, so additional experiments were designed to test this hypothesis. Previous activity assays, undertaken in the laboratory, using cell extracts included a competitor oligo in the reaction mix, to order to prevent degradation of [α-^{32}P]-labelled substrate molecules by cellular nuclease. Bovine serum albumin (BSA) was also included in the assay buffer, to help stabilise TDP2. The presence of BSA and the competitor oligo abrogated the inhibitory effect of C14005 (Figure 4.5). After removing

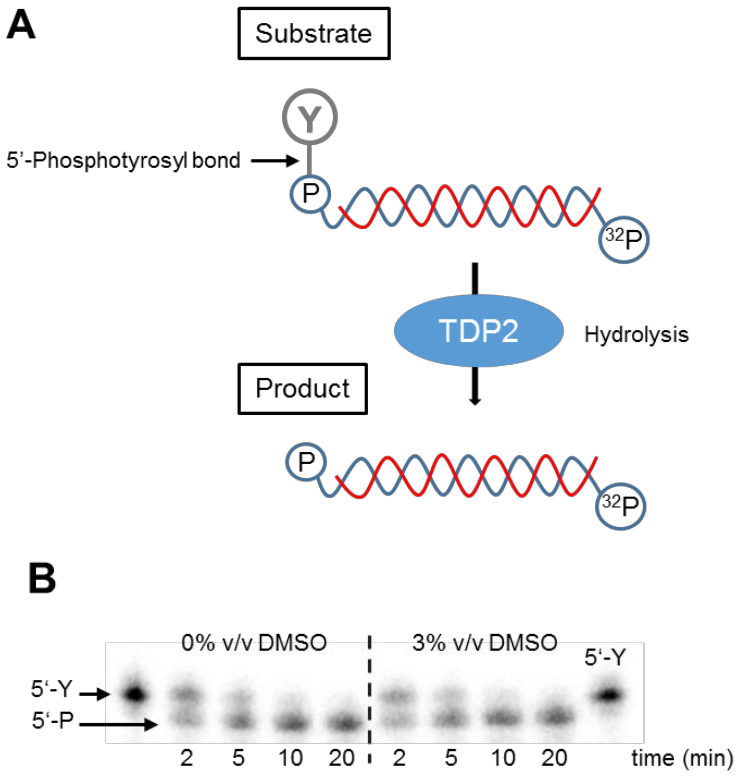

Figure 4.4. Gel-based enzyme activity assay for hTDP2-CAT^{C161S} in the presence 3% v/v DMSO. **A.** Schematic representation of [α-^{32}P]-labelled oligonucleotide duplexes with a 5' overhang (4 nt) harbouring a 5'-phosphotyrosine moiety. **B.** Gel-based assay to test the impact of 3% v/v DMSO on enzyme activity. Bands corresponding to the substrate 5'-Y-oligo and the product 5'-P-oligo were resolved by denaturing sequencing gel.

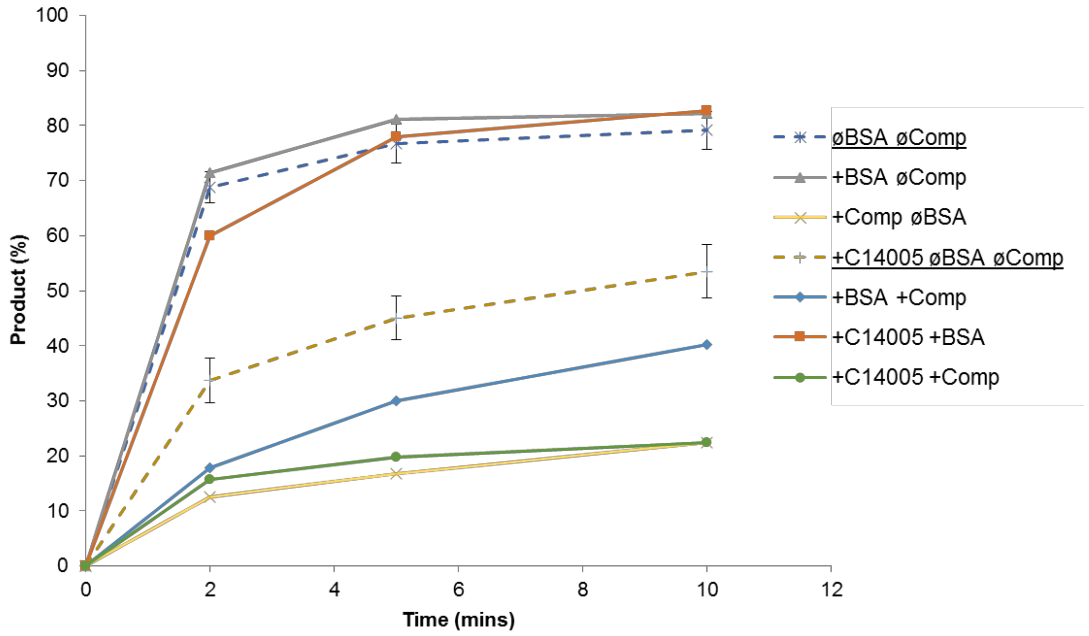

Figure 4.5. Quantification of gel-based activity assays for hTDP2-CAT^{C161S} in the presence of compound C14005, testing different experimental conditions. The essay was conducted in the presence of 1nM hTDP2-CAT$^{161-406}$ and 50nM inhibitory compound (C14005) and at 3% v/v DMSO. KEY: BSA: bovine serum albumin, Comp: competitor DNA oligo. Conditions where BSA and competitor oligo were excluded from the reaction, thus allowing an inhibitory effect of compound C14005 to be seen, are marked with dashed lines to aid visualisation.

these two reagents from the experiment, a significant inhibitory effect of C14005 on hTDP2 could be observed (Figure 4.6).

4.2.3 Co-crystallization trials of hTDP2-CAT^{C161S} and C14005

As both the thermal denaturation and activity assays indicated that C14005 bound to and inhibited hTDP2, co-crystallisation trials were undertaken (Materials and Methods 2.4.1.2). Crystals of hTDP2^{C161S} in complex with C14005, were obtained in the following condition: 100mM bis-tris propane pH7, 0.45M NaCl, 0.05$MgAc_2$, 1.5% v/v trimethylamine *N*-oxide (TMAO), supplemented with 3% v/v DMSO to improve compound solubility. However, the low solubility of C14005 did not allow a molar ratio of protein: compound greater than 1:1. Greater ratios would have required a DMSO concentration of >3%. Although DMSO is often used to aid protein crystallisation, at higher concentrations it can promote protein precipitation and denaturation (Arakawa et al. 2007).

Crystals typically grew after five days at 20ºC (Figure 4.7). They were harvested, cryo-protected by successive swipes through buffer containing increasing concentrations of sucrose (final 40% v/v), and then plunge-frozen in liquid N_2 before being stored. Diffraction data was collected on beamline I04, DLS. Unfortunately, these crystals only diffracted to low resolution (8-9Å).

In order to obtain a higher resolution dataset, a compound with greater solubility was required, in order to improve the protein: compound molar ratio to greater than 1:2, and to keep the DMSO concentration relatively low (≤ 3% v/v). Consequently, through discussions with the Paterson Institute, a more soluble tetrazole compound (C14297) (Figure 4.8) with a similar sub-micromolar disassociation constant was provided.

4.2.4 Crystallization of hTDP2 in complex with compound C14297

Thermal denaturation assay of hTDP2-CAT^{C161S} in the presence of C14297 showed a robust positive shift in Tm (ΔTm=3°C) (Figure 4.9). Consequently, co-crystallisation trials with hTDP2 in complex with C14297 were set up, at an increased protein:compound molar ratio of 1 to 2.7 (due to the improved solubility of the tetrazole compound) (Materials and Methods 2.4.1.2). Crystals were obtained with the SaltRx HT screen (Hampton Research) in MRC2 sitting-drop plates in condition C9 (1.2M DL Malic acid pH 7.0, 0.1M bis-tris propane pH7.0) after 5 days of incubation at 20ºC. These crystals diffracted fairly poorly

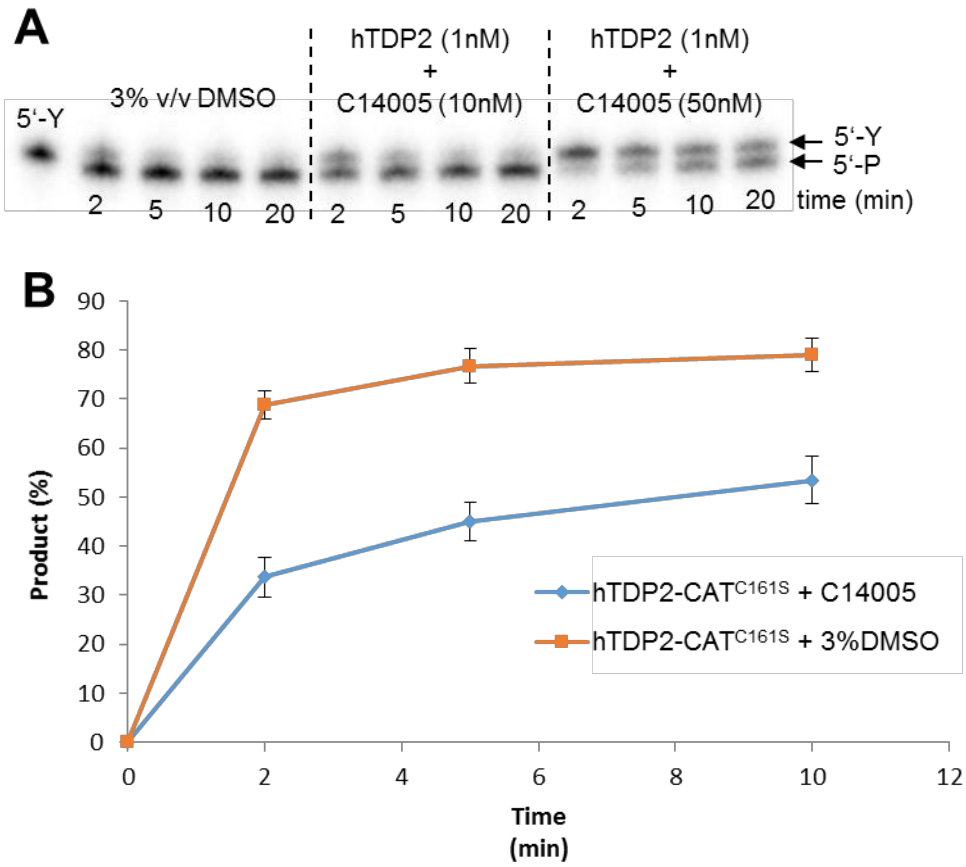

Figure 4.6. Gel-based enzyme activity assay in the presence of compound C14005. **A.** Reactions were conducted in the presence of 1nM hTDP2-CAT$^{161\text{-}406}$, and 10 and 50nM inhibitory compound (C14005), and then resolved on denaturing sequencing gel. **B.** Quantification of the inhibitory effect of C14005 on hTDP2 enzyme activity. Data are the mean of three independent experiments, with error bars representing one standard deviation.

Figure 4.7. Representative image of hTDP2-CAT^{C161S} / C14005 co-crystals. Crystallisation condition: 100mM bis-tris propane pH7.0, 0.45M NaCl, 0.05$MgAc_2$, 1.5% v/v trimethylamine *N*-oxide (TMAO), 3% v/v DMSO.

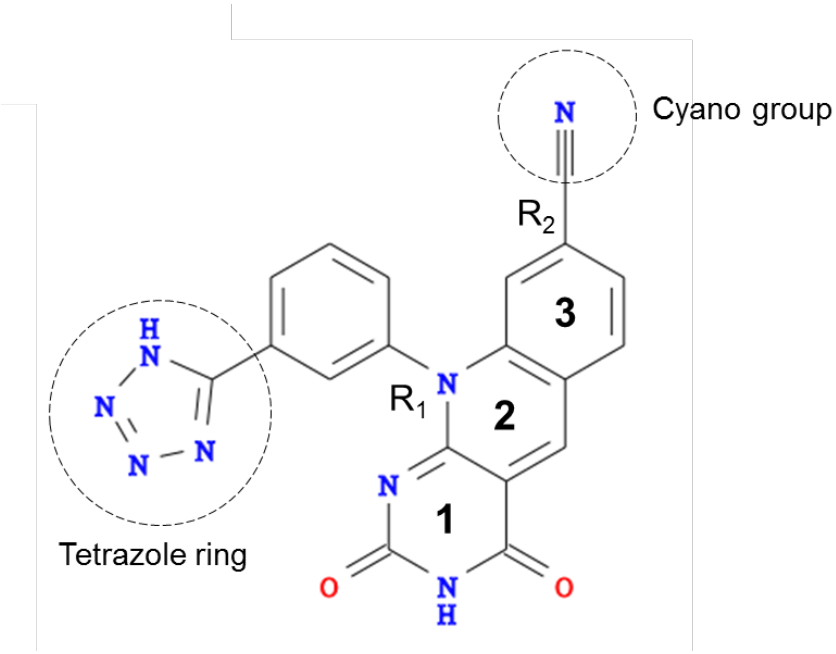

Figure 4.8. Chemical drawing of C14297. The compound is a tetrazole derivative (R_1 substituent) of the deazaflavin series with an EC_{50} value in the submicromolar range (50nM). Numbers for the three rings of the deazaflavin core are also shown.

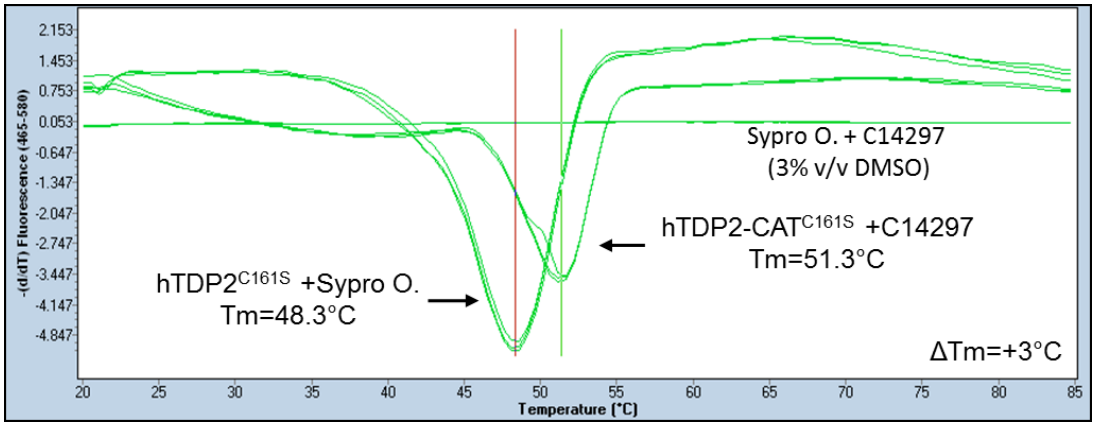

Figure 4.9. Thermal denaturation assay of hTDP2-CAT^{C161S} in the presence of compound C14297. First order derivatives of the observed melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Final compound concentration: 300µM. Reaction conditions: HEPES pH7.0, 300mM NaCl, 0.5mM TCEP and 3% v/v DMSO.

and anisotropically, but by screening several crystals it was possible to collect an isotropic dataset on beamline I04, DLS to a resolution of 3.1Å. No further optimisation of crystallisation conditions was undertaken.

4.2.5 Molecular Replacement

At this time, the structure of the catalytic domain from murine TDP2 (mTDP2-CAT) was published (PDB: 4GYZ, (Schellenberg et al. 2012)) at a resolution of 2.4Å. It was therefore possible to solve by molecular replacement the structure of both the apo- and liganded forms of $hTDP2^{C161S}$, using the atomic coordinates of the murine protein. Molecular replacement used PHASER (CCP4 software suite, (Winn et al. 2011)). The initial molecular replacement model was then improved by iterative rounds of rebuilding and refinement, with COOT (Emsley et al. 2010) and REFMAC5 (Vagin et al. 2004) to obtain the final refined model.

The catalytic domain of human TDP2 has the same fold as other phosphodiesterases of the Mg-dependent phosphodiesterase class (e.g. AP endonuclease). It is comprised of two central β-sheets surrounded by eight alpha helices (Figure 4.10A).

Electron density maps clearly indicated the position of the bound compound. Surprisingly, it did not bind directly to the enzyme's active site (Figure 4.11); rather it made contact with the large hydrophobic platform formed by the 'M7-loop' of the enzyme (Figure 4.12A). The M7- loop was originally described by Schellenberg et al. (2012). It consists of two highly conserved short amino acid motifs 'TWDT' and 'RFDR' that are separated by a poorly conserved 14 amino acid linker (Figure 4.12B), and serves to generate a platform for interaction with an incoming DNA substrate (Schellenberg et al. 2012) (Figure 4.11B).

As the resolution of the diffraction data was relatively low, only a few conclusions could be drawn from the modelled structure. In particular, the Nε of the Arg266 side chain appeared to be poised to form a hydrogen bond to the carbonyl group located on ring 1 of the deazaflavin ring system. Several amino acid residues from the divergent region of the M7-loop make stacking interactions with the core of the compound. (Figure 4.13). Arg231, which as part of the M5 'cap' motif, helps bind an incoming DNA substrate by interacting with the sugar-phosphate backbone, also appears to play a role in compound binding and orientation (Schellenberg et al. 2012).

At low resolution, the placing of water molecules becomes difficult and problematic; as only electron density for highly coordinated water molecules is generally seen. The

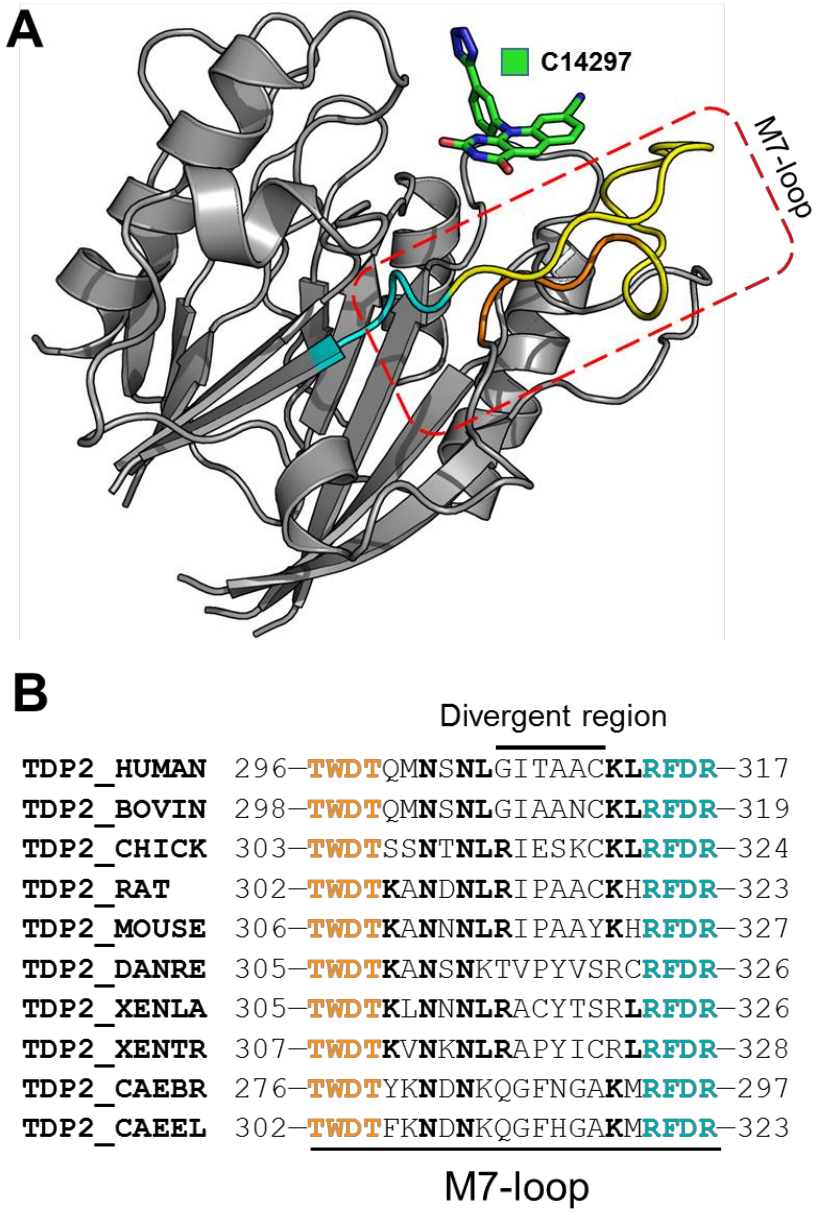

Figure 4.10. Cartoon representation for the structure of hTDP2-CAT$^{161\text{-}406}$ in complex with compound C14297. **A.** Schematic representation of the structure of the catalytic domain of hTDP2 in complex with the compound highlighting the M7-loop. The positions of the conserved 'TWDT' and 'RFDR' amino acid sequence motifs, which flank the M7-loop, are respectively indicated by the cyan and orange-coloured regions. Residues of the divergent M7-loop, are additionally highlighted in yellow, and by the dotted line red coloured box. **B.** Multiple amino acid sequence alignment for the M7-loop across several different species indicates the divergent region.

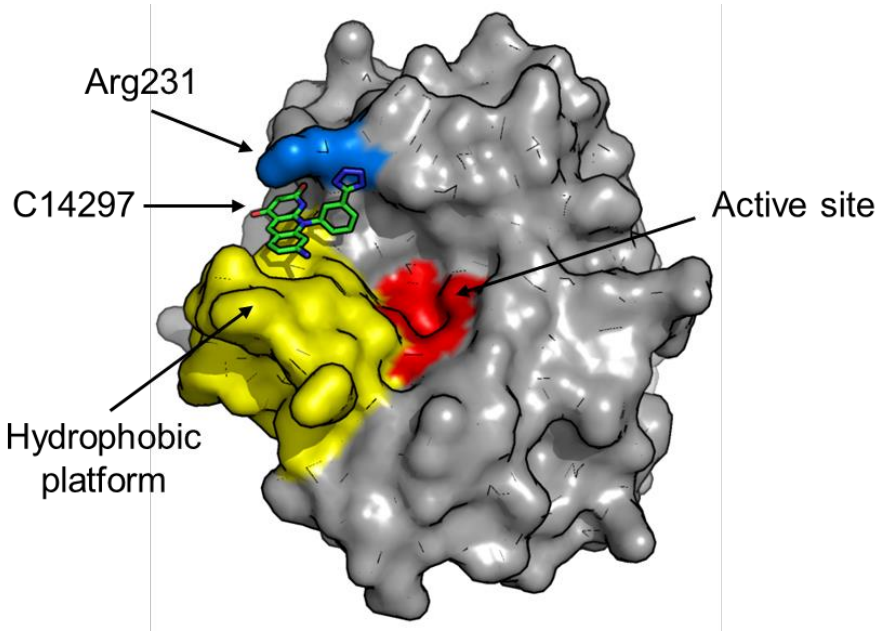

Figure 4.11. Molecular surface representation of hTDP2-CAT^{C161S} highlighting the position of C14297 relative to the enzyme's active site. The active site and hydrophobic platform are coloured red and yellow, respectively. For orientation, the sidechain of Arg231 is additionally highlighted in blue.

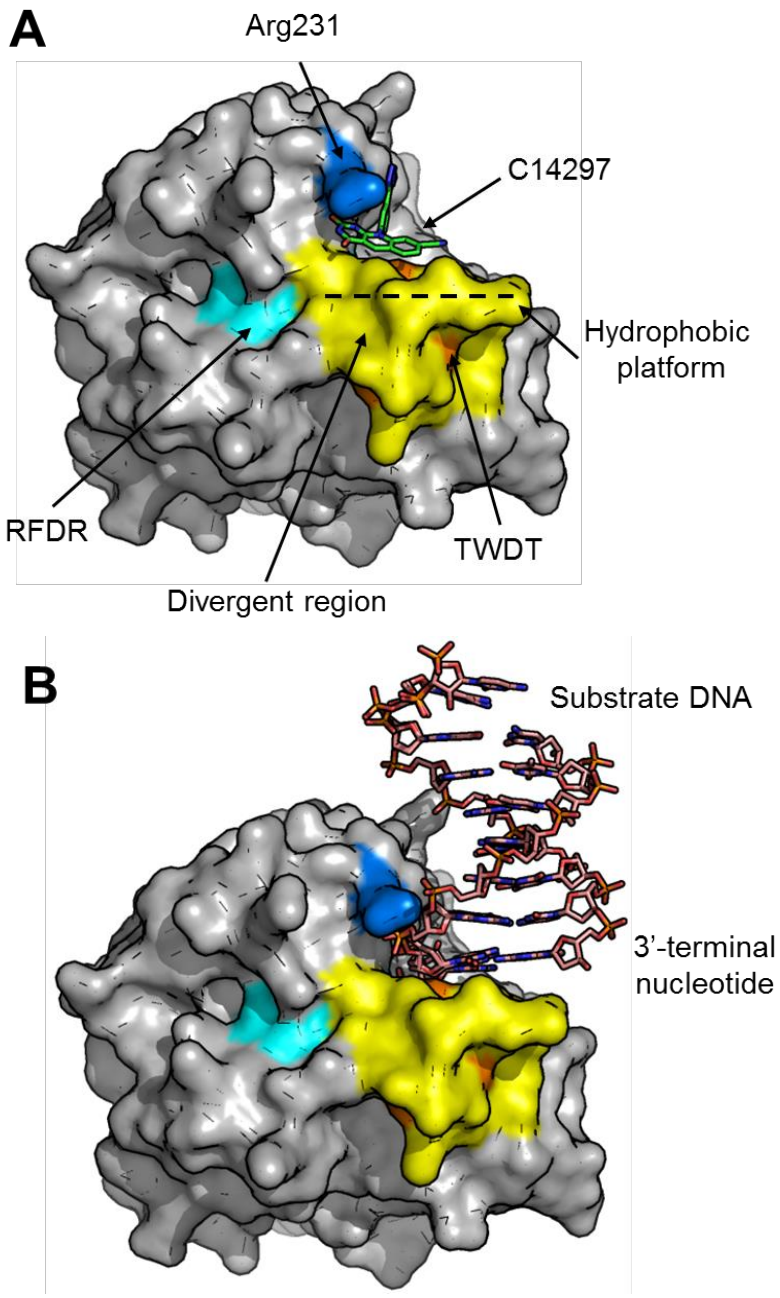

Figure 4.12. C14297 binds to the 'hydrophobic platform' of hTDP2. **A.** Molecular surface representation for hTDP2-CAT^{C161S} in complex with C14297. The hydrophobic platform, formed by residues of the M7-loop, making several van der Waals interactions with the inhibitory compound. Regions of the M7-loop highlighted similar to Figure 4.10. Arg231 -part of the DNA binding motif 'cap'- is highlighted in blue. **B.** Figure depicts the positioning of the substrate DNA by the DNA binding motifs.

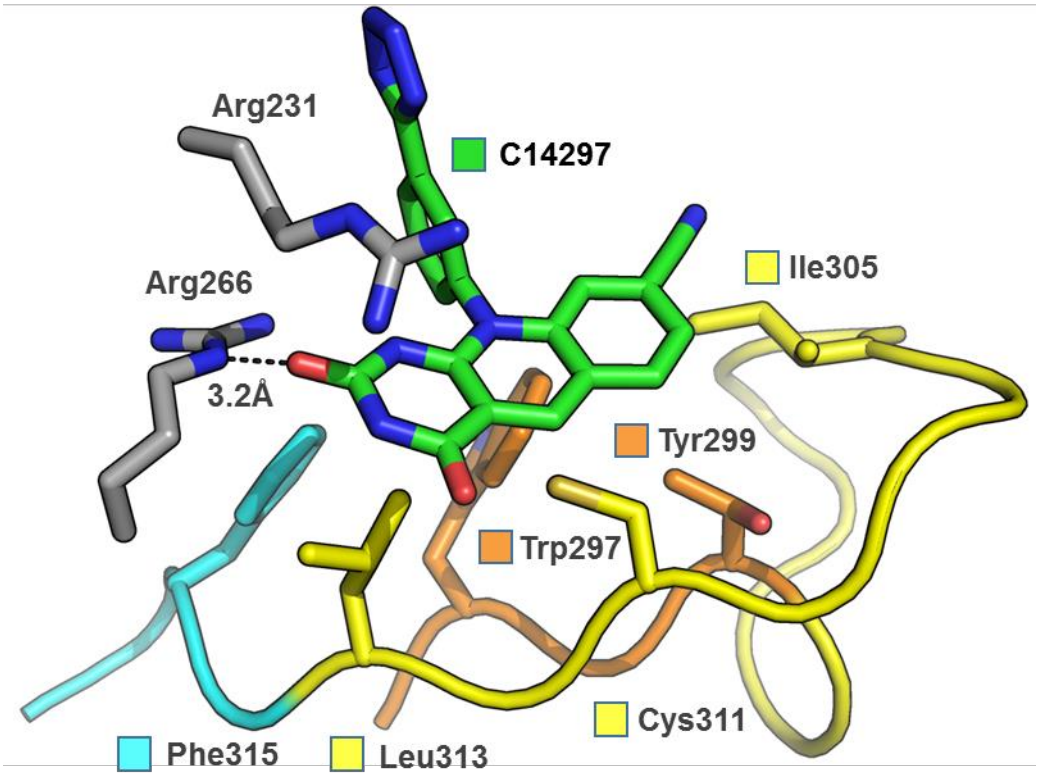

Figure 4.13. Schematic figure, indicating key amino acids that are involved in C14297 binding. Compound binding is facilitated via hydrogen bond formation and stacking interactions. The DNA binding platform formed by the M7-loop is highlighted by orange (TWDT motif), yellow (divergent region), and cyan (RFDR motif).

position of water molecules can often provide valuable information for rational drug design, by 'mapping' potential hydrogen bond donors/acceptors, and indicating a potential 'path' for expansion of a compound. Ordered water molecules can often be replaced by water mimicking substituents, through chemical expansion of the inhibitory compound, to increase both its binding affinity and specificity (Lounnas et al. 2013).

As crystals of hTDP2-CAT^{C161S} typically diffracted poorly and anisotropically, we sought a better system for our drug discovery efforts. As the structure of the catalytic domain from mouse TDP2 was solved at high resolution, it was a logical step to employ this protein in structural studies.

4.2.6 Thermal denaturation and activity assays with mTDP2-CAT

4.2.6.1 Thermal denaturation assay with C14297

Expression and purification of mTDP2-CAT is described in section 2.2.6.5. The mTDP2-CAT expression construct contains an N-terminal polyhistidine-tagged SUMO moiety that is cleavable by Sentrin-specific protease 1 (SENP1) (Figure 4.14). Following purification (Figure 4.15 and 4.16), mTDP2-CAT was used in thermal denaturation assays in order to verify binding of C14297 to the enzyme (Figure 4.17) (Material and Methods 2.6.1). However, there was no observable shift in melting temperature in these experiments, suggesting that the compound did not bind the murine form of the enzyme.

4.2.6.2 Enzyme activity assays with C14297

To verify that C14297 did not bind or inhibit the murine form of TDP2, it was taken forward into the gel-based assay (Materials and Methods 2.3.2). There was a significant difference in the inhibition profiles of the two enzymes; the compound being far more active on the human protein that on the mouse (Figure 4.18). It is important to note that the enzyme activity of the recombinant mTDP2-CAT was approximately 3-fold less than that of the hTDP2-CAT^{C161S}. In order to obtain comparable data of the inhibition of both enzymes by the compound, the experiments were conducted in the presence of 1nM hTDP2-CAT^{C161S} and 3nM mTDP2-CAT. Together these experiments confirmed that C14297 poorly bound and inhibited the murine protein.

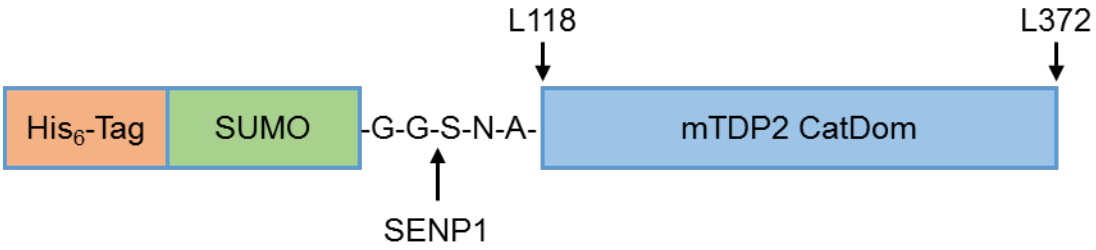

Figure 4.14. Schematic representation of the murine TDP2-CAT expression construct. The construct encodes the catalytic domain of the catalytic domain (amino acid boundaries Leu118 and Leu372) and an N-terminal His-SUMO solubility tag, in order to improve yields of soluble recombinant protein. The tag can be removed by overnight digestion by SENP1 protease.

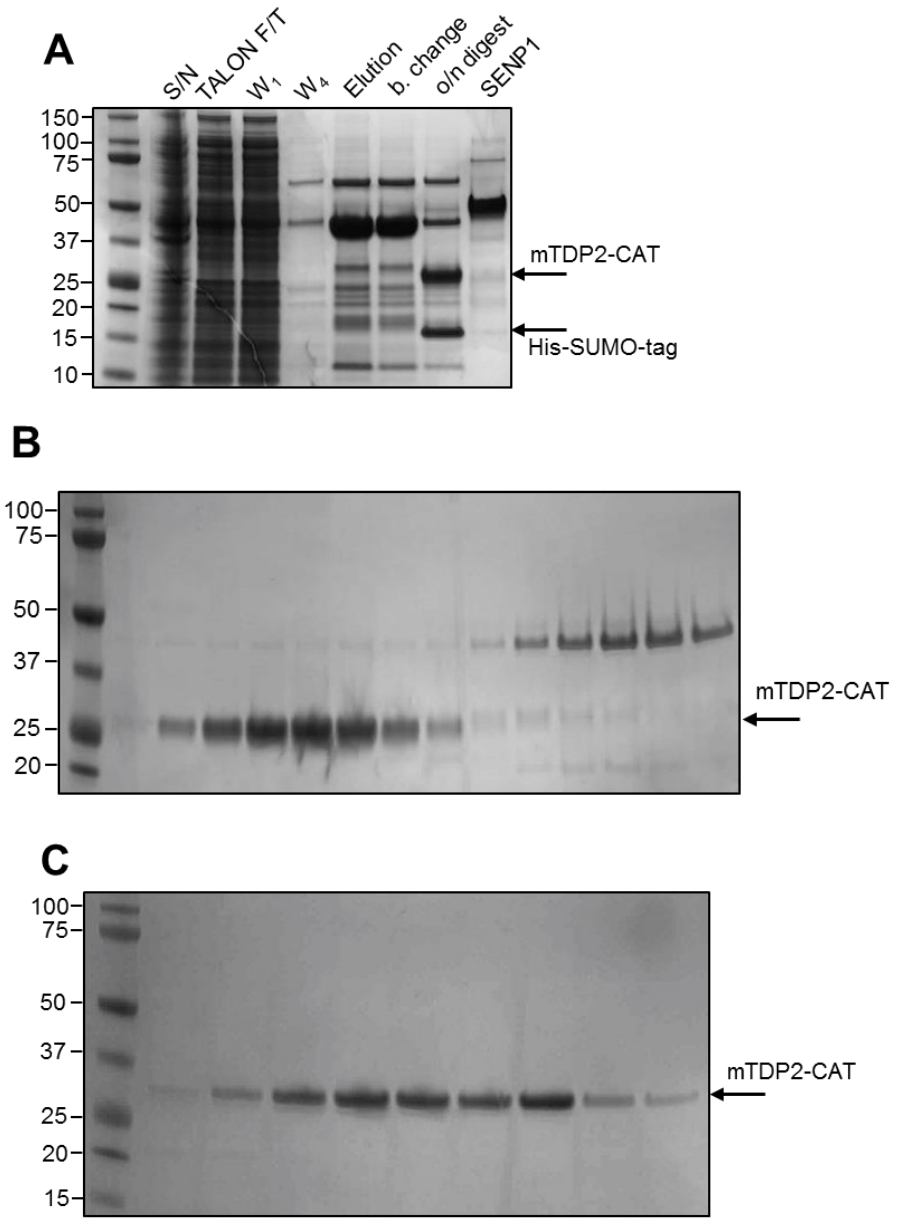

Figure 4.15. Representative images of SDS-PAGE gels for key steps in the purification of murine TDP2 (mTDP2-CAT). **A.** Clarified supernatant (S/N) was loaded onto TALON resin. The eluent was subjected to buffer exchange (b. change) before it was digested with SENP1 overnight. **B.** The overnight digest was loaded onto Heparin column and eluted with increasing NaCl. **C.** Eluent from Heparin column was further purified by gel filtration column. Samples were run on Novex 4-12% Tris-Glycine Protein Gels and then stained with InstantBlue Protein Stain. KEY: (S/N): supernatant, (F/T): flow-through, (W_1/W_2): wash, (b. change): buffer change, (o/n digest): overnight digest. An arrow indicates the expected size of the recombinant protein and the removed His-SUMO-tag. MW of mTDP2-CAT is 28.8 kDa.

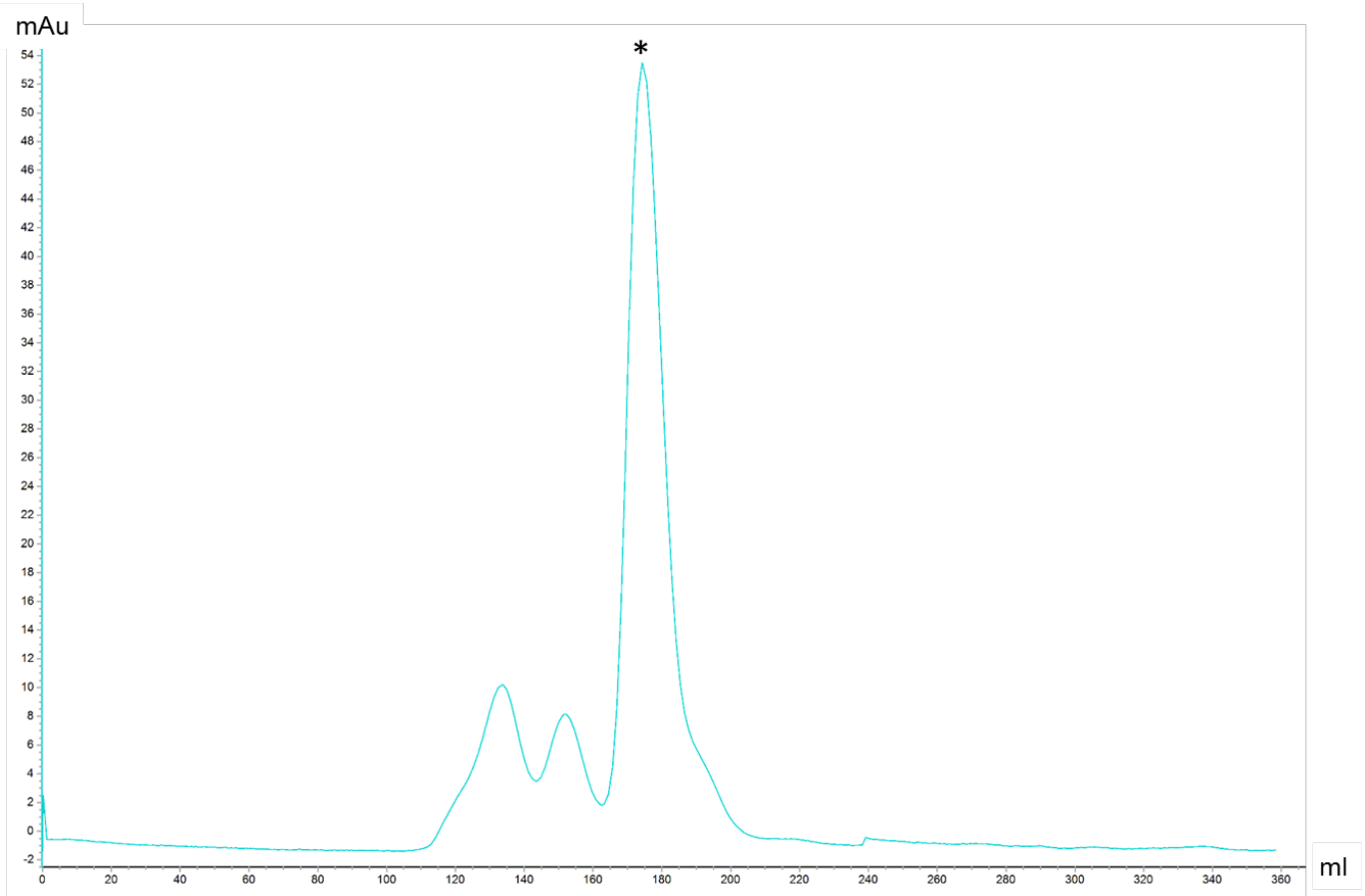

Figure 4.16. Gel filtration profile of m2hTDP2. Following elution from Heparin column with increasing NaCl the sample was further purified by Superdex S75 gel filtration column (GE Healthcare).

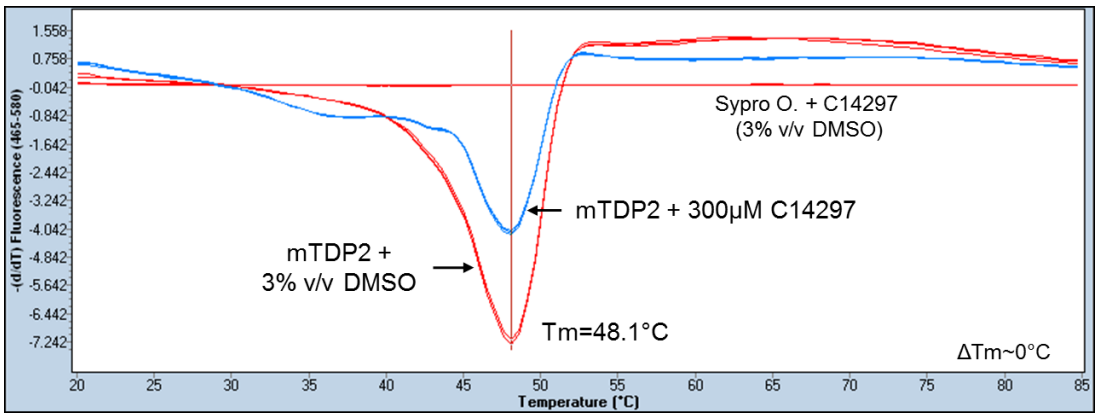

Figure 4.17. Thermal denaturation assay of the catalytic domain of murine TDP2 in the presence of compound C14297. First order derivatives of the observed melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Final compound concentration: 300µM. Reaction conditions: HEPES pH7.0, 300mM NaCl, 0.5mM TCEP and 3% v/v DMSO.

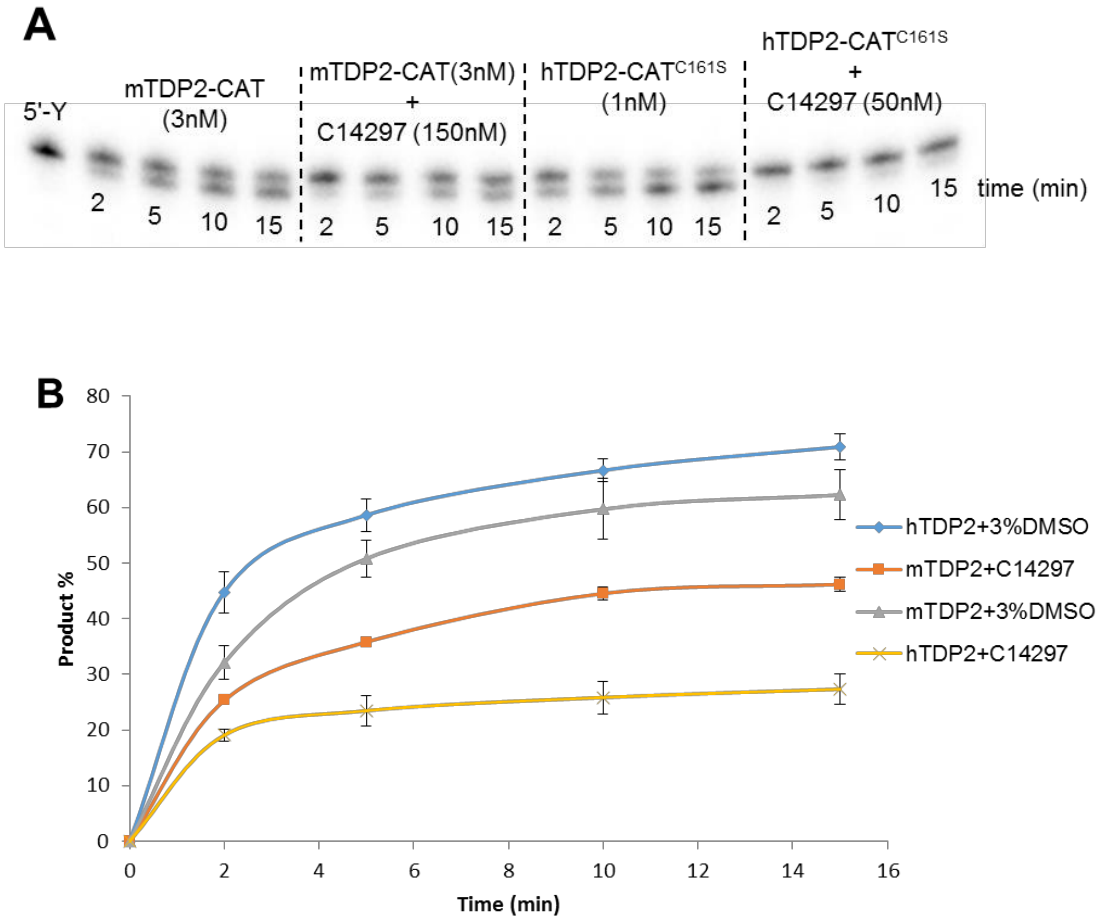

Figure 4.18. Activity assays comparing the inhibitory effect of compound C14297 on hTDP2-CAT and mTDP2-CAT. **A.** Reactions were conducted in the presence of 1nM hTDP2-CAT$^{161\text{-}406}$ and 3nM mTDP2-CAT at 3% v/v DMSO (mTDP2-CAT was approximately 3 fold less active than hTDP2-CAT^{C161S}). To compare the inhibitory effect of C14297 on enzyme activity, human and murine TDP2 were incubated in the presence of 50nM and 150nM inhibitory compound, respectively, and then resolved on denaturing sequencing gel. **B.** Quantification of the inhibitory effect of C14297 on hTDP2 and mTDP2. Data are the mean of three independent experiments, with error bars representing one standard deviation.

4.2.7 Structure comparisons of the catalytic domains of human and mouse TDP2

To try to explain the observed differences in inhibition profiles, we compared the structure of the catalytic domain of mTDP2 (Schellenberg et al. 2012) to hTDP2 (Figure 4.19). hTDP2-Cys311 is replaced by a tyrosine residue in the mouse protein (mTDP2-Tyr321); through model comparisons it is evident that Tyr321 is likely to occlude compound binding to mTDP2. Moreover, this residue is located within the divergent region of the M7-loop (Figure 4.10B) and its identity is not conserved across species. Furthermore, other non-conserved residues (including those of the M7-loop) are critical for a number of protein-compound interactions (Figure 4.13, Figure 4.19). Three other amino acid substitutions were thought to play significant roles in abrogating compound binding. 1) Leu313 forms an integral part of the 'hydrophobic platform' and its alteration to histidine in mouse is likely to disrupt compound binding. 2) Gly232, which is changed to a glutamate in the mouse, may produce steric clashes with the tetrazole ring of the compound. 3) Gln278, which is altered to arginine in the mouse, will affect the hydrogen-bonding potential for any bound compound. For more a detailed understanding of the interaction between hTDP2 and compound C14297, crystals that diffracted isotropically to higher resolution were required. As the murine protein seemed to provide such a system, it was hypothesised that by mutation of each of the amino acids described above, back to their human counterparts, compound binding might be restored. This work is described in the following Chapter.

4.3 Discussion

Our primary goal was to obtain structural information about the catalytic domain of human TDP2 in order to assist a structure-based drug design approach, for the development of small-molecule inhibitors of TDP2. A series of deazaflavin compounds were provided by the Cancer Research UK Drug Discovery Unit at the Patterson Institute for Cancer Research (University of Manchester). In order to find suitable candidates for structural studies, the compounds first were subjected to binding assays.

The thermal denaturation assay has become a commonly employed technique, not only for optimising buffer conditions to promote protein stability, but also for screening compounds for binding (Lo et al. 2004, Senisterra et al. 2012). By use of this assay format, it was

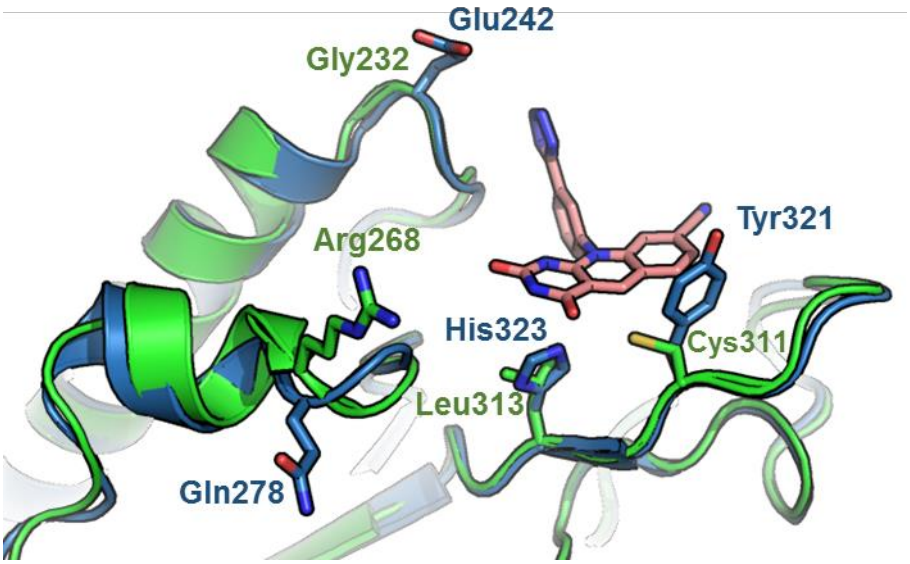

Figure 4.19. Molecular cartoon representation for hTDP2-CAT^{C161S} aligned with mTDP2. Carbon atoms of hTDP2-CAT^{C161S} and mTDP2 are coloured in green and blue, respectively. The four amino acids mutated to create m2hTDP2-CAT are shown in stick representation. C14297 is also show in stick representation (carbon atoms coloured in coral), to highlight the position of bound compound, and its relation to the mutated amino acids.

possible to show that several deazaflavin compounds bound to the catalytic domain of hTDP2, producing a positive shift in temperature midpoint (Tm). In particular, compound C14005, a phenylmethanesulfonamide derivative of the deazaflavin series, produced the largest increment in melting temperature (3.5 °C).

As C14005 clearly bound and stabilised hTDP2, it was taken forward into enzyme activity assays, to examine any inhibitory affect. The assay used is a modified form of that reported by Ledesma *et al*., who determined that TDP2 is a 5'-tyrosyl DNA phosphodiesterase (Cortes Ledesma et al. 2009). Whilst that assay was performed using cell extracts, in the modified format, the activity of recombinant protein is tested. Initially, there was no observable inhibition of C14005 on the enzyme activity of human TDP2. However, by removing competitor oligo and bovine serum albumin (BSA) from the reaction mix, it was possible to restore and monitor the inhibitory effect of the compound. Competitor oligo was originally included in order to prevent degradation of [α-^{32}P]-labelled substrate molecules by contaminating nucleases, whereas BSA stabilized TDP2 activity in cell lysates.

These results suggest that C14005 is likely to be a weakly binding inhibitor of TDP2; the presence of competitor oligonucleotide at high concentration (400μM) prevented compound binding to hTDP2, presumably by preventing access to its binding site by direct competition. The presence of BSA has a similar effect, again preventing compound from binding to TDP2, presumably by sequestering the compound due to non-specific binding and/or aggregation effects. However, as dissociation constants have not been determined for each compound/protein interaction, this hypothesis requires confirmation. Future experimental work should therefore ideally include determination of binding constants for each compound, using a suitable technique such as iso-thermal titration calorimetry (ITC).

Nevertheless, in the modified assay, C14005 clearly inhibited the enzymatic activity of recombinant hTDP2. So, in order to determine the molecular parameters for how (and where) the compound binds the protein, the complex was taken into crystallisation trials. However, the poor aqueous solubility of C14005 hindered this approach. However, a new opportunity presented itself when an additional compound (C14297) was provided that had improved solubility characteristics. Crystals were obtained for this compound in complex with hTDP2-CAT^{C161S} and diffraction data collected. It was possible to obtain phases by

molecular replacement, using the atomic coordinates of the catalytic domain of murine TDP2 as a search model. The resultant electron density maps clearly indicated the position and binding mode of C14297. Perhaps surprisingly, the compound did not bind directly to the active site, rather it made a series of van der Waals contacts with the so-called 'hydrophobic platform', formed by the M7-loop first described by Schellenberg et al., (Schellenberg et al. 2012) and providing an interaction surface for the incoming DNA substrate. However, as crystals of hTDP2-CAT^{C161S} typically diffracted anisotropically and to low resolution, and alternative strategy was sought. It was initially thought that the murine version of the enzyme (mTDP2-CAT) might provide a viable alternative. However, thermal denaturation and gel-based activity assays demonstrated that the compound did not demonstrably bind, or inhibit mTDP2 to the same or similar extent as it did with hTDP2.

Structural comparisons of the catalytic domains of mTDP2 and hTDP2 show that there are certain amino acid changes, in close proximity to the compound binding site that may explain the lack of an inhibitory effect. In order to substantiate the hypothesis that these particular amino acid changes (or at least some of them) inhibited compound binding, a 'humanised' version of the murine catalytic domain was created, in which the residues were mutated 'back' to their human counterparts; i.e. E242G, Q278R, Y321C and H323L. It was anticipated that this 'humanised' construct (called m2hTDP2-CAT, 'mouse-to-human') would provide co-crystal structures at higher resolution, and without anisotropic diffraction. This experimental work is described in the following chapter.

CHAPTER FIVE

Structures of ‘humanised’ murine TDP2 in complex with small molecule inhibitors

5.1 Introduction

In the previous chapter, the structure of the catalytic domain of human TDP2 in complex with a small molecular inhibitor, provided preliminary information about the binding mode of the deazaflavin series. However, the resolution of diffraction data resulting from hTDP2-CAT^{C161S} crystals was limited and was typically highly anisotropic. Therefore, in order to obtain higher resolution structures, a protein 'surrogate' of human TDP2 was developed: m2hTDP2-CAT. In this expression construct, four amino acids of the murine protein, lining the compound-binding pocket, were mutated to their human counterpart in order to restore binding and enzyme inhibition by the deazaflavin compounds.

5.2 Results

5.2.1 Expression and purification of m2hTDP2-CAT

The expression and purification protocol for the humanised catalytic domain of mouse TDP2 (m2hTDP2-CAT), is described in Materials and Methods (2.2.6.6). This expression construct encodes an N-terminal polyhistidine-tagged SUMO moiety that is cleavable by Sentrin-specific protease 1 (SENP1) (Figure 5.1). During cell lysis and Talon capture steps (i.e. the first purification steps) a HEPES buffer at a pH of 7.5 was used. However, after overnight digestion with SENP1 protease –to remove the His-SUMO solubility tag – the theoretical isoelectric point (pI) of the m2hTDP2-CAT protein drops from pH6.2 to pH7.5. As a ThermoFluor assay clearly indicated that lowering the pH greatly increased the thermal stability of the protein (Figure 5.2), a MES buffer at a pH of 6.0 was used during subsequent chromatographic steps, which served to keep the recombinant protein in solution (Figure 5.3).

5.2.2 Enzyme activity assays

Gel-based enzyme activity assays were performed with purified m2hTDP2-CAT, in parallel with hTDP2-CAT^{C161S} and mTDP2-CAT as controls (Material and Methods 2.3.2). As predicted, the activity of m2hTDP2-CAT was similar to both the murine and the human proteins. More importantly, C14297 now demonstrably inhibited m2hTDP2-CAT to a similar extent as hTDP2-CAT, thereby validating our hypothesis and approach (Figure 5.4).

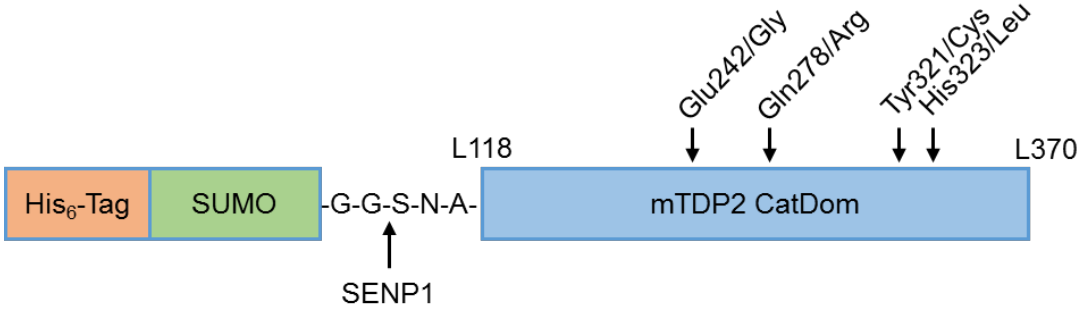

Figure 5.1. Schematic representation of the 'humanised' catalytic domain of mouse TDP2 (m2hTDP2-CAT). The amino acids mutated to their human counterpart are indicated. The expression construct encodes an N-terminal His-SUMO affinity/solubility tag to improve yields of soluble recombinant protein. The tag can be removed by digestion with SENP1 protease.

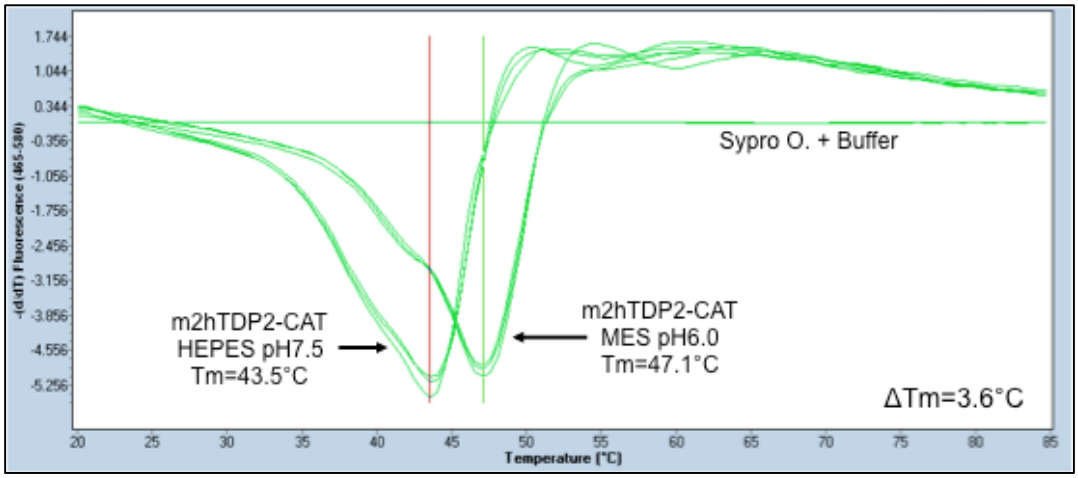

Figure 5.2. Thermal denaturation assay: m2hTDP2-CAT in different buffer systems. First order derivatives of the measured melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Reaction conditions: HEPES pH7.5 (or MES pH6.0), 300mM NaCl, 0.5mM TCEP.

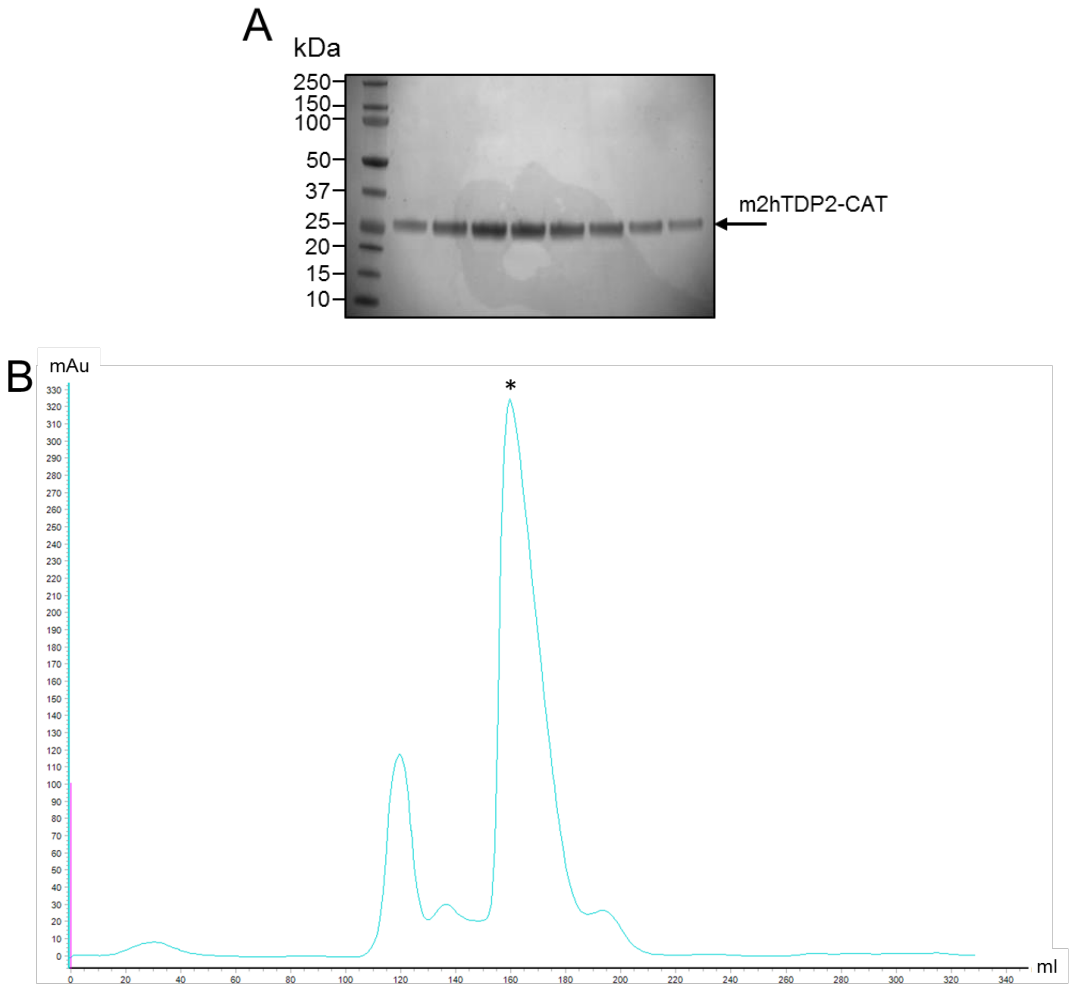

Figure 5.3. Purification of recombinant m2hTDP2. **A.** A representative SDS-PAGE gel, for fractions eluted from a HiLoad 16/600 Superdex 75pg size exclusion chromatography column across the indicated peak (asterisk). The expected migration position of the recombinant protein is indicated by an arrow. **B.** A representative chromatograph for the size exclusion chromatography purification step.

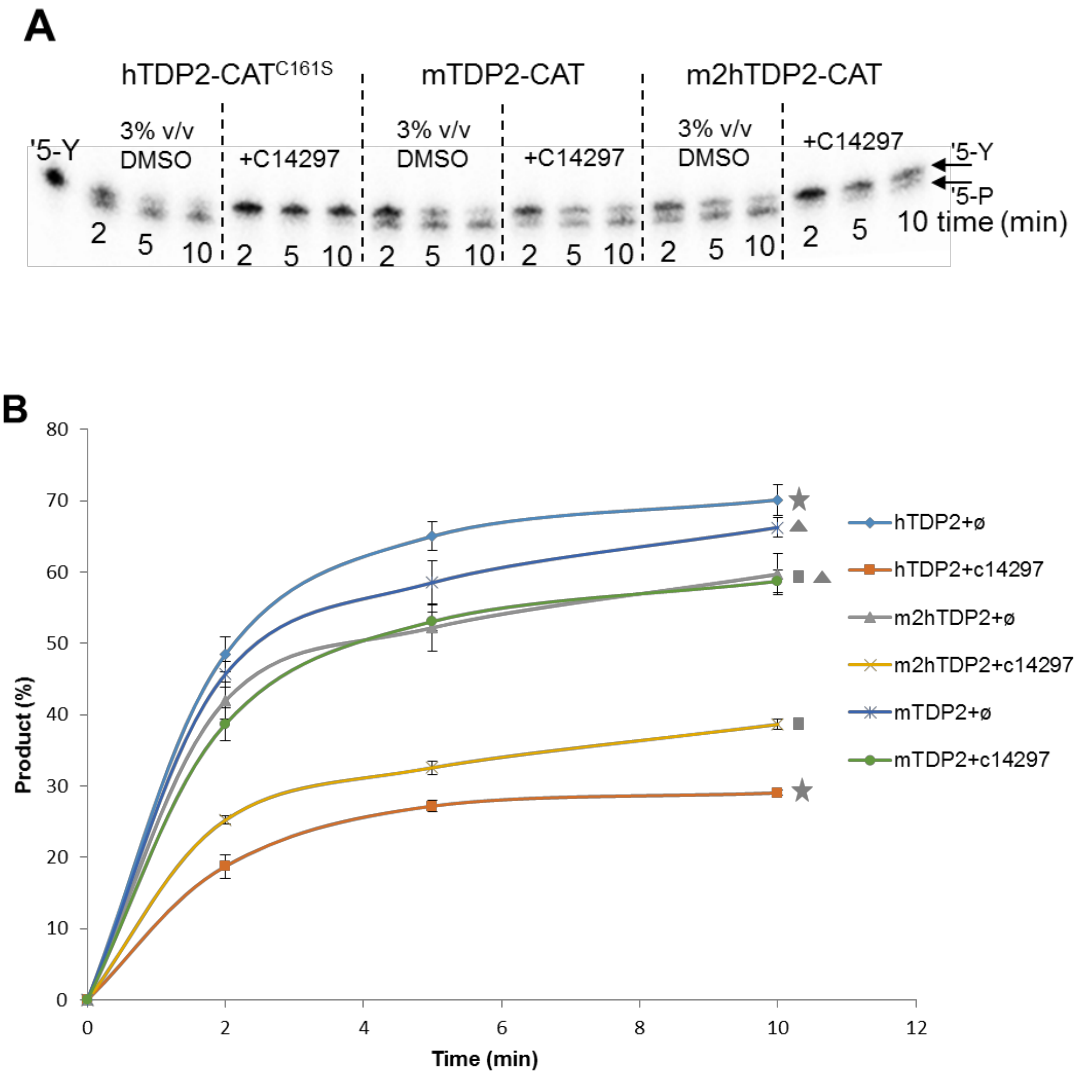

Figure 5.4. Gel-based enzyme activity assay, comparing the inhibitory effect of compound C14297 on all three recombinant TDP2 proteins. **A.** Reactions were conducted with 1nM hTDP2-CAT^{C161S}, 3nM mTDP2-CAT and 1nM m2hTDP2-CAT, in the presence of 50, 150 and 50nM C14297, respectively (for an explanation of the different enzyme concentrations please see Materials and Methods 2.3.2) and then resolved on a denaturing sequencing gel. **B.** Quantification of the inhibitory effect of C14297 on enzyme activity. Plotted data are the mean of three independent experiments. Error bars represent 1 standard deviation. The star, triangle and square symbols correspond to hTDP2, mTDP2 and m2hTDP2 activity, respectively, for easier comparison.

5.2.3 Crystallization of m2hTDP2-CAT in complex with C14297

Crystallisation trials, in MRC2 sitting drop plates, were set up as before (Materials and Methods 2.4.1.4) using several commercially available screens. After 5 days of incubation at 20°C, crystals from condition G11 of the PACT *premier* HT-96 screen (0.2M Sodium citrate, 0.1M bis-Tris propane pH7.5, 20% w/v PEG 3350, Molecular Dimensions) were harvested (Figure 5.5) and cryo-protected by successive swipes through buffer containing increasing concentrations of cryo-protectant (8% v/v glycerol and 5% w/v glucose) and then plunge-frozen in liquid N_2 before being stored. Diffraction data to a resolution of 1.7Å were collected from a single crystal, on beamline I04, DLS. As crystals readily diffracted to high resolution, no further optimisation was required (Table 5.1). The dataset was auto-processed at Diamond through the xia2 system using the 3dii mode (Winter, 2010), and then indexed and integrated by Mosfilm (Leslie, 2006) and scaled by XSCALE (Kabsch, 2010).

5.2.4 Molecular replacement

Molecular replacement, with the published structure of mTDP2-CAT as a search model (PDB: 4GYZ) in PHASER (CCP4 software suite (Winn et al. 2011)) was sufficient to phase the diffraction data. The molecular replacement model was then improved by iterative rounds of rebuilding and refinement, using COOT (Emsley et al. 2010) and REFMAC5 (Vagin et al. 2004) to obtain the final refined model. Water molecules were placed in position where the sigma level was 2.0 or above and positioned in relation to hydrogen bonding to the protein or other ligands. Models were validated with the MolProbity server (Chen et al. 2010) at several stages throughout building, to inspect and correct side-chain rotamers, and to highlight problematic region, i.e. Ramachandran outliers. The final model had an R-factor of 0.1954 (R_{free} 0.2256) and agreed well with electron density (Table 5.1).

Two molecules of m2hTDP2-CAT comprise the asymmetric unit (Figure 5.6). Interestingly, they are covalently coupled by a disulphide bond formed between the two protomers via the side chain of cysteine 137. It is likely that this disulphide bond is created opportunistically during the exposure of crystals to X-rays, as m2hTDP-CAT behaves as a monomer in solution, under most experimental conditions (unpublished observations).

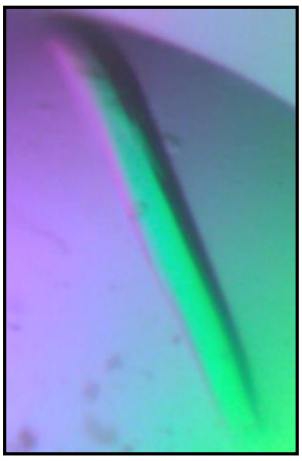

Figure 5.5. Image of a m2hTDP2-CAT / C14297 co-crystal. Crystals emerged from condition G11 (0.2M Sodium citrate, 0.1M bis-tris propane pH7.5, 20% w/v PEG 3350) from PACT *premier* HT-96 screen (Molecular Dimensions) after 5 days of incubation at 20°C.

Data collection statistics	
Wavelength (Å)	0.97949
Unit cell parameters	
a (Å)	61.130
b (Å)	42.450
c (Å)	108.680
α (°)	90.000
β (°)	93.950
γ (°)	90.000
Space group	P 1 2_1 1
Resolution range	54.79-1.46 (1.50-1.46)*
Completeness	99.3 (98.1)
Multiplicity	3.0 (2.8)
Mean(*I*)/σ(*I*)	19.9 (2.1)
Unique reflections	96352 (1159)
R_{merge}	0.028 (0.432)
R_{pim}	0.025 (0.350)
Refinement statistics	
Protein molecules per in the asymmetric unit	2
R_{work} / R_{free}	0.1954 / 0.2256
Number of atoms	4549
Protein	3927
Ligands	76
Water	546
Protein residues	494
Ramachandran plot (%)	
Favoured	98
Additionally allowed	2
RMSDs	
Bond lengths (Å)	0.01
Bond angles (°)	1.46
B-factor (Å^2)	
Average B-factor	25.90
macromolecules	23.11
ligands	33.90
solvent	45.10

Table 5.1. Data collection and refinement statistics for the m2hTDP2-CAT / C14297 complex. *Values in parentheses are for highest-resolution shell.

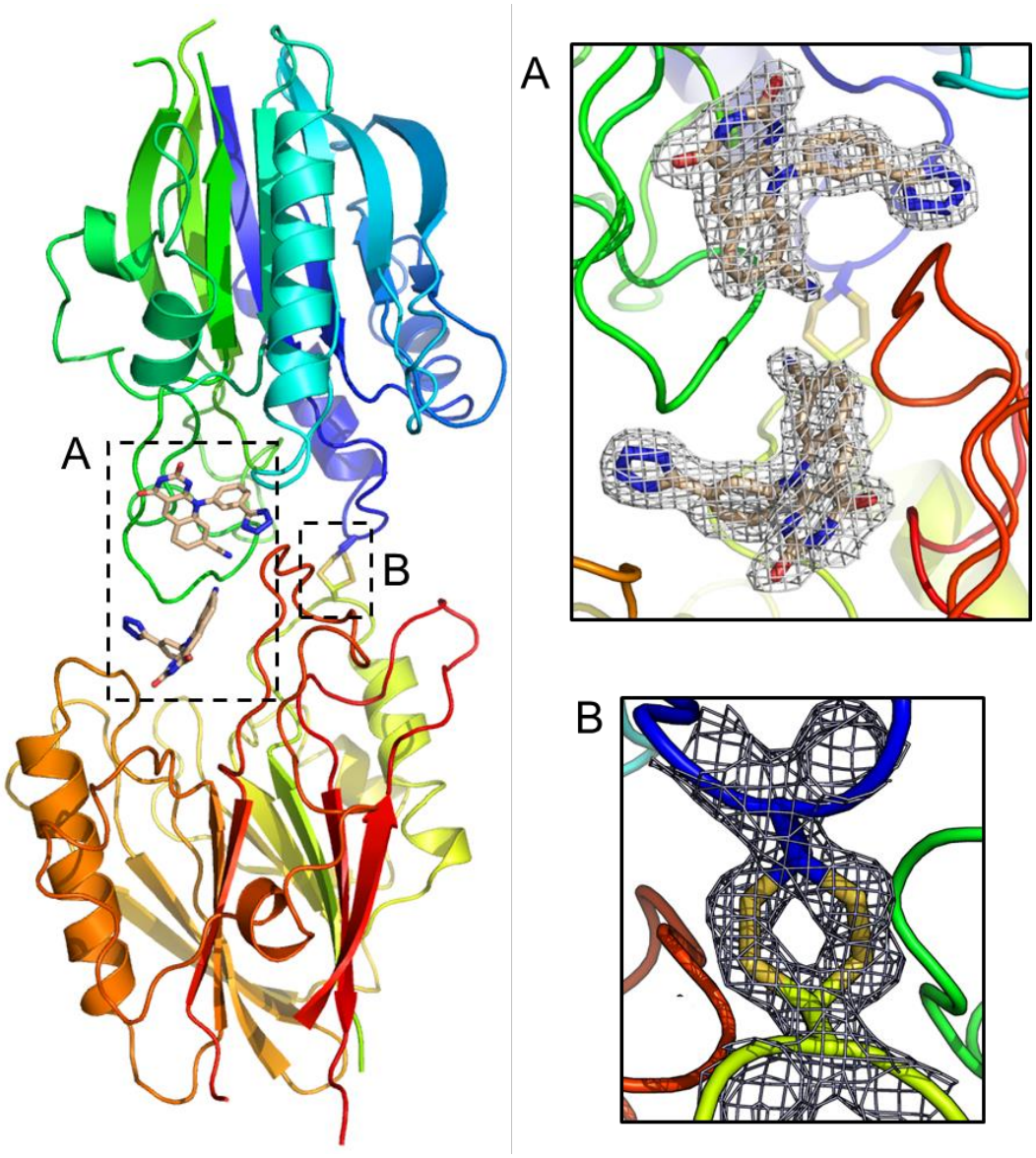

Figure 5.6. Molecular cartoon representation of the asymmetric unit of m2hTDP2-CAT crystals. **A.** C14297 bound to each molecule of the asymmetric unit is shown in stick representation (carbon atoms coloured brown). Inset: Representative $2F_o$-F_c electron density map (grey mesh, contoured at 1.5σ) surrounding both C14297 molecules **B.** Disulphide bond formed (Cys137 to Cys137) between the two protomers of the asymmetric unit. Inset: Representative $2F_o$-F_c electron density map for the disulphide bond; note: two alternative conformational states are modelled, each at half occupancy.

5.2.5 Structure of m2hTDP2-CAT in complex with C14297

As described for human TDP2 (Chapter 4.2.5), the deazaflavin ring system of bound C14297 – through a series of van der Waals contacts – interacts with residues from the divergent region of the M7-loop, which include Trp307, Ile317, Ala319, Cys321 and Leu323. However, due to the improved resolution, the positions of many ordered water molecules could now be observed. As before, the guanidinium head group (atoms Nε and Nη) of Arg266 (Arg276 in hTDP2-CAT^{C161S}) is poised to form hydrogen bonds to the carbonyl and nitrogen atoms of ring 1 of the deazaflavin core (Figure 5.7); Arg266 also interacts with the side chain of Asn274 via a water molecule. The 'back-mutated' Arg268 (Arg278 in hTDP2-CAT^{C161S}) also makes contact with the same ring system via a water-mediated interaction. The side chain of Arg241 (Arg251 in hTDP2-CAT^{C161S}) also helps to stabilise compound binding by stacking with both the pyrimidodione moiety (ring 1) and the tetrazole substituent. It is worth noting here, that due to its close proximity, Pro318 from the other protomer of the asymmetric unit is involved in an intermolecular stacking interaction with the benzene moiety and tetrazole ring of C14297. However, as this interaction most likely emerges from the fact that the two m2hTDP2-CAT proteins are covalently connected via a disulphide bond, it is likely to be a crystallisation artefact. A ligand interaction diagram for m2hTDP2-CAT in complex with C14297 is presented in Figure 5.8.

5.2.6 Crystallization of m2hTDP2-CAT in complex with LEP-0259

As part of an ongoing collaboration with the Sussex Drug Discovery Centre, a new series of inhibitory compounds, based on the deazaflavin series, were synthesised. These compounds had higher aqueous solubility, with similar EC_{50} values (in the nM range) to the earlier deazaflavin compounds. In order to assess their binding capability, as before, each of the compounds was first tested in the thermal denaturation assay. Of the 4 compounds tested, 3 did not show any measureable shift in temperature midpoint (ΔTm); however, one compound, a 3-hydroxyphenol derivative (LEP-0259, Figure 5.9), produced a small, but reproducible shift (ΔTm=1.3ºC) (Figure 5.10). This compound was therefore taken forward into crystallisation trials.

In this instance, during purification of the m2hTDP2-CAT protein, the gel filtration buffer was supplemented with manganese chloride (iso-structural with magnesium, but readily discernible in electron density maps) in order to more fully define the catalytic environment

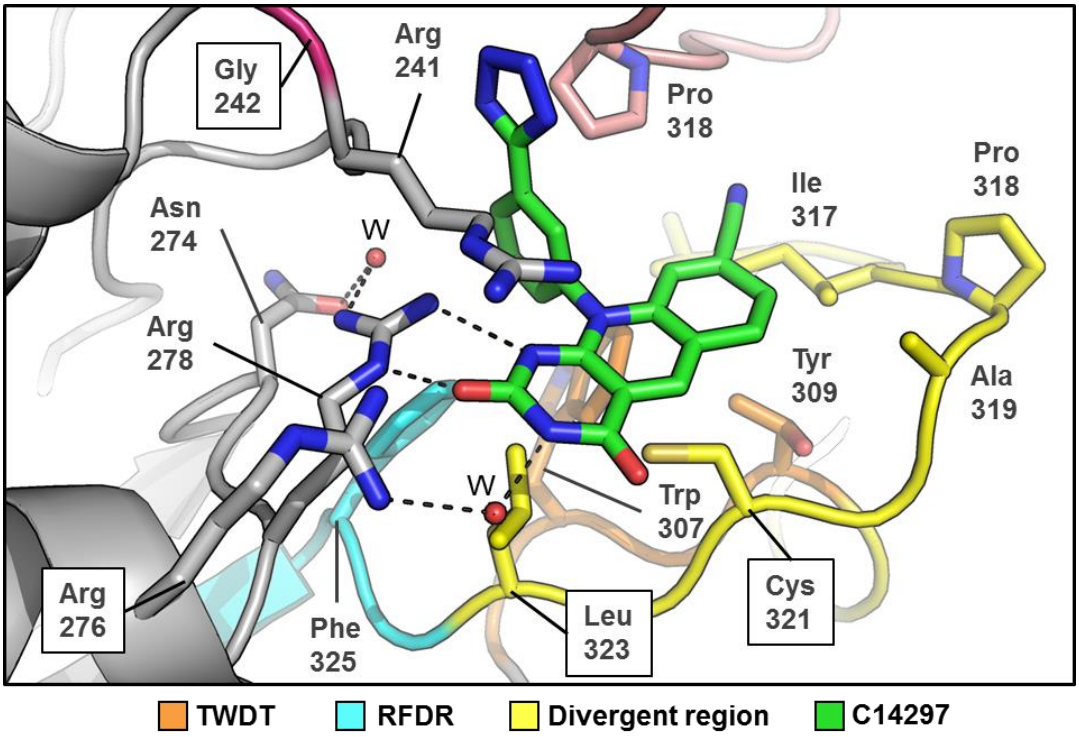

Figure 5.7. Molecular cartoon representation of m2hTDP2-CAT in complex with C14297, highlighting the residues surrounding the compound-binding site. The 'hydrophobic platform' formed by residues of the M7-loop is shown, and is represented by amino acids coloured in orange (TWDT motif), yellow (divergent region) and cyan (RFDR motif). Pro318, which contacts C14297, from the other protomer of the asymmetric unit, is coloured in salmon. C14297 is shown in stick representation, and is coloured in green (carbon atoms). Dotted lines represent potential hydrogen bonds. Water molecules involved in compound interactions are shown as red spheres, and labelled with the letter 'w'.

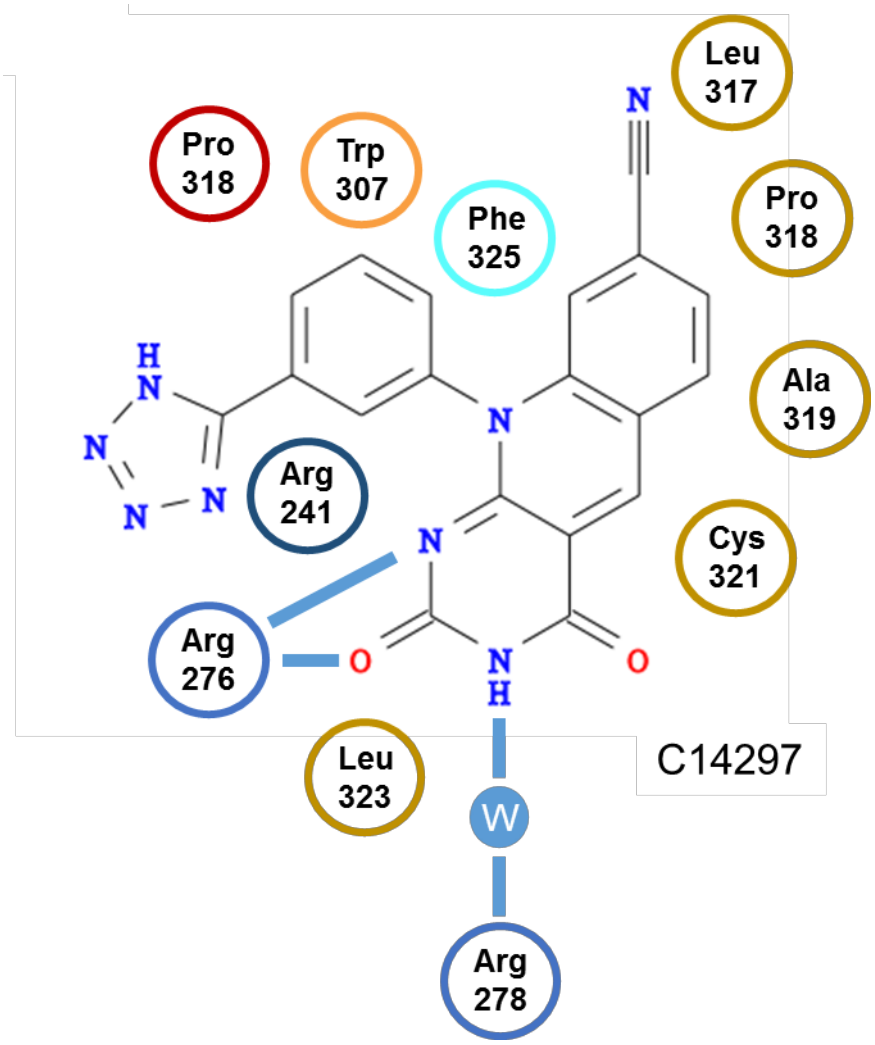

Figure 5.8. C14297 ligand interaction diagram. Key: W: water; blue arrows represent hydrogen bonds, dashed line: coordination bond. Blue circles represent residues involved in hydrogen bond formation and water mediated interaction, the orange circle represents Trp308 from the TWDT motif, the cyan circle depicts Phe325 from the RFDR motif, yellow circles represent residues from the divergent region. Pro318 represented as a red circle from the other protomer in the asymmetric unit is also involved in stacking interaction.

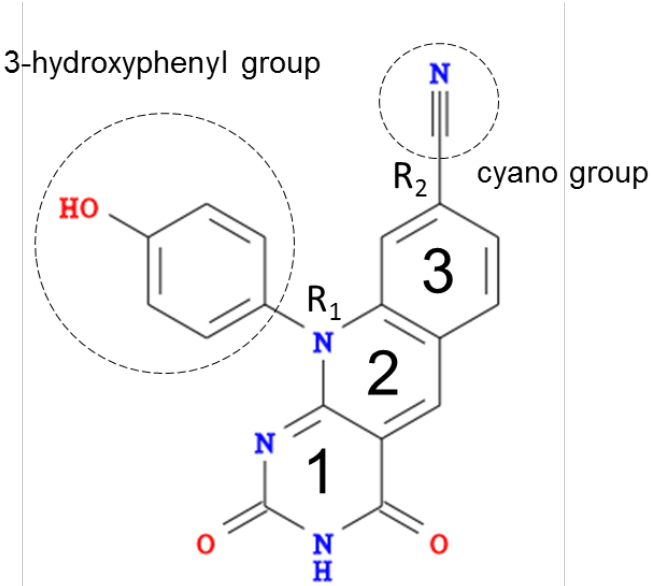

Figure 5.9. Chemical drawing of LEP-0259. The compound is a 3-hydroxyphenol derivative (R_1 substituent) of the deazaflavin series with an EC_{50} value in the submicromolar range (50nM). Numbers for the three rings of the deazaflavin core are also shown.

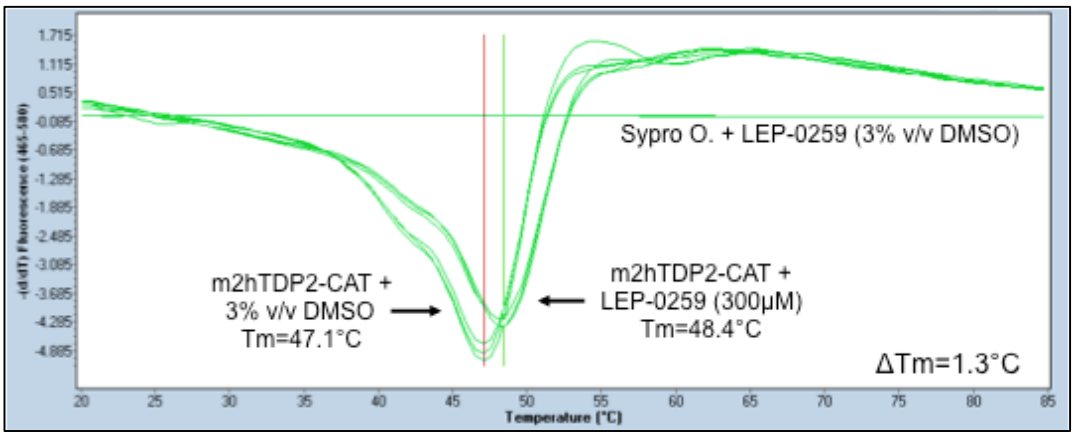

Figure 5.10. Thermal denaturation assay of m2hTDP2-CAT in the presence of LEP-0259. First order derivatives of the measured melting curves are shown (rate of change of fluorescence intensity). Temperature midpoints for the transition from folded to unfolded states are determined by the indicated peak minima (vertical line) in each case. Protein concentration: 1.7µM. Final compound concentration: 300µM. Reaction conditions: 20mM MES pH6.0, 300mM NaCl, 0.5mM TCEP and 3% v/v DMSO.

of the enzyme. As TDP2 is a magnesium-dependent phosphodiesterase, Mn^{2+} was expected to form a coordination complex at the active site of the enzyme.

Crystallisation trials, in MRC2 sitting drop plates, were set up as before (Materials and Methods 2.4.1.5) using several commercially available screens. After 5 days of incubation at 20°C, crystals from condition F9 of the PACT *premier* HT-96 screen (0.2M Potassium sodium tartrate tetrahydrate, 0.1M bis-Tris propane pH6.5, 20% w/v PEG 3350; Molecular Dimensions) were harvested (Figure 5.11), and cryo-protected by successive swipes through buffer containing increasing concentrations of cryo-protectant (8% v/v glycerol and 5% w/v glucose) and then plunge-frozen in liquid N_2 before being stored. Diffraction data to a resolution of 1.6Å were collected from a single crystal on beamline I04, DLS,

As the space-group was identical, and the cell-dimensions were similar to m2hTDP2-CAT / C14297 crystals, a simple rigid-body refinement was sufficient to phase the diffraction data (Table 5.2).

Again the starting model was then improved by iterative rounds of rebuilding and refinement, using COOT and REFMAC5 as before. Again, models were validated with the MolProbity server at several stages throughout building (Chen, Arendall et al. 2010).

5.2.7 Structure of m2hTDP2-CAT in complex with LEP-0259

As in the previous complex, two TDP2 molecules comprised the asymmetric unit, again connected together by the same disulphide bond, formed by the side chain of Cys137 in each protomer.

The binding mode of LEP-0259 is essentially identical to that observed for C14297 (Figure 5.12). Residue Arg241, part of the 'cap' motif described by Schellenberg et al, is packed directly against the face of the tricyclic core. As before, the guanidinium group of Arg276 forms several hydrogen bonds with the carbonyl and nitrogen atoms of ring 1 of the deazaflavin core of the compound, and also involved in a water-mediated interaction with the carbonyl group of Asn274. Arg278 stabilizes compound binding further through an ordered water molecule.

Interestingly, the hydroxyl group of the R_1 substituent in the *para* position, points directly towards the catalytic core of the protein – and makes a series of water-mediated contacts to the side chains of Asp272 and Asn274. Electron density maps also clearly showed the

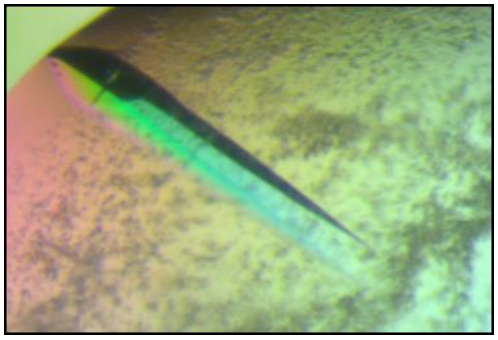

Figure 5.11. Image of a m2hTDP2-CAT / LEP-0259 co-crystal. Crystals emerged from condition F9 (0.2M Potassium sodium tartrate tetrahydrate, 0.1M bis-tris propane pH6.5, 20% w/v PEG 3350) from PACT *premier* HT-96 (Molecular Dimension) screen after 5 days of incubation at 20°C.

Data collection statistics	
Wavelength (Å)	0.97949
Unit cell dimensions:	
a (Å)	61.040
b (Å)	42.860
c (Å)	109.010
α (°)	90.000
β (°)	94.050
γ (°)	90.000
Space group	P 1 2_1 1
Resolution range (Å)	54.80-1.65 (1.69-1.65)*
Completeness (%)	99.5 (98.6)
Multiplicity	3.1 (2.8)
Mean(*I*)/σ(*I*)	17.7 (2.1)
Unique reflections	67784 (2371)
R_{merge}	0.04 (0.432)
R_{pim}	0.035 (0.367)
Refinement statistics	
Protein molecules per in the asymmetric unit	2
R_{work} / R_{free}	0.2054 / 0.2361
Number of atoms	4526
Protein	3894
Ligands	68
Manganese	2
Tartrate	4
Water	522
Protein residues	490
Ramachandran plot (%)	
Favoured	97.5
Additionally allowed	2.5
RMSDs	
Bond lengths (Å)	0.015
Bond angles (°)	1.51
B-factor (Å^2)	
Average B-factor	24.40
macromolecules	25.11
ligands	34.70
solvent	47.10

Table 5.2. Data collection and refinement statistics for m2hTDP2-CAT / LEP-0259 complex. *Values in parentheses are for highest-resolution shell.

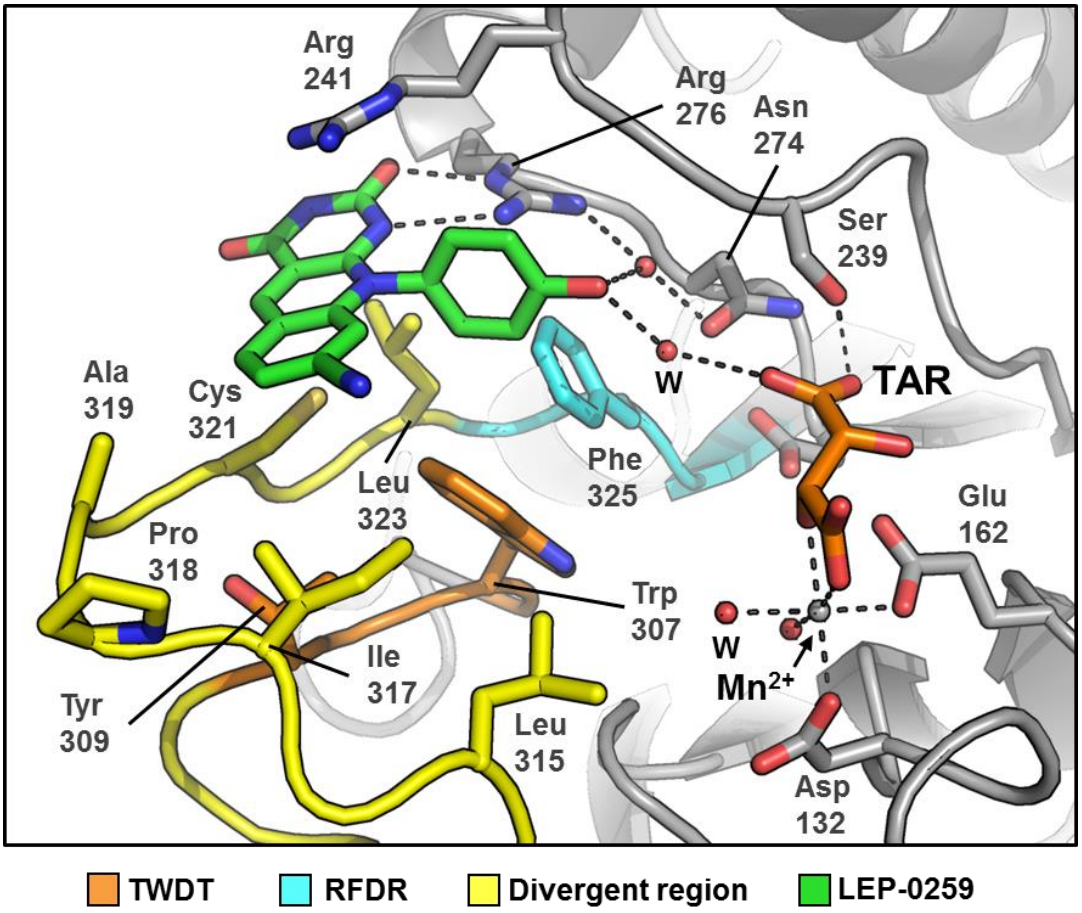

Figure 5.12. Molecular cartoon representation of m2hTDP2-CAT in complex with LEP-259, highlighting the residues surrounding the compound-binding site. The 'hydrophobic platform' formed by residues of the M7-loop is shown, and is represented by amino acids coloured in orange (TWDT motif), yellow (divergent region) and cyan (RFDR motif). LEP-0259 is shown in stick representation, and is coloured in green (carbon atoms). Dotted lines represent potential hydrogen bonds. Water molecules involved in compound interactions are shown as red spheres, and labelled with the letter 'w'. KEY: TAR: tartrate, Mn^{2+}: manganese.

position of a tartrate molecule (picked up from the crystallisation solution) which is hydrogen-bonded to the side chain of Ser239, as well as an extensive series of water-mediated interactions with several amino acid residues surrounding the catalytic centre. The tartrate molecule also completes the octahedral coordination of the Mn^{2+} ion (marking the position of the catalytic Mg^{2+} co-factor).

A ligand interaction diagram for m2hTDP2-CAT in complex with compound LEP-0259 is presented in Figure 5.13.

5.3 Discussion

In this chapter, two high-resolution structures of m2hTDP2-CAT in complex with small molecule inhibitors are described.

Several derivatives of deazaflavin-based compounds have been proven to be effective *in vitro* inhibitors of TDP2, with reported EC_{50} values in the sub-micromolar range (Raoof et al. 2013). Molecular docking studies, predicted that this series of compounds bound to TDP2 via ring 1 (pyrimidodione ring) of the tricyclic core, and that bound directly to the magnesium co-factor that is required for TDP2 enzyme activity, and made little (if any) hydrogen bonds directly to the protein (Raoof et al. 2013). However, the x-ray crystal structures of m2hTDP2-CAT in complex with two different inhibitors, presented here, show that the docking poses are in fact incorrect, and the binding of deazaflavins is actually driven by the tricyclic core making a series of van der Waals contacts to the hydrophobic platform of TDP2 (formed by the M7-loop). In particular, the hydroxyl and nitrogen atoms of ring 1, of the deazaflavin core, are liganded by the side chain of Arg278. Residues Arg241 and Cys321 also provide additional contacts through stacking interaction with the top and bottom of the tri-cyclic ring system.

The tetrazole ring of C14297 was thought to increase compound solubility (Raoof et al. 2013) but was originally introduced to provide a chemical group that would provide additional hydrogen-bonding potential, if presented towards the catalytic core of TDP2. However, early docking studies did not indicate any specific interactions, and it was hypothesised that the tetrazole ring might form non-specific hydrogen bonds and water mediated interactions with polar residues either in the hydrophobic DNA-binding 'wall' or on the opposite loop, due to the *meta* substitution of the tetrazole moiety (Raoof et al. 2013). However, in the hTDP2-CAT / C14297 complex, the tetrazole ring is actually

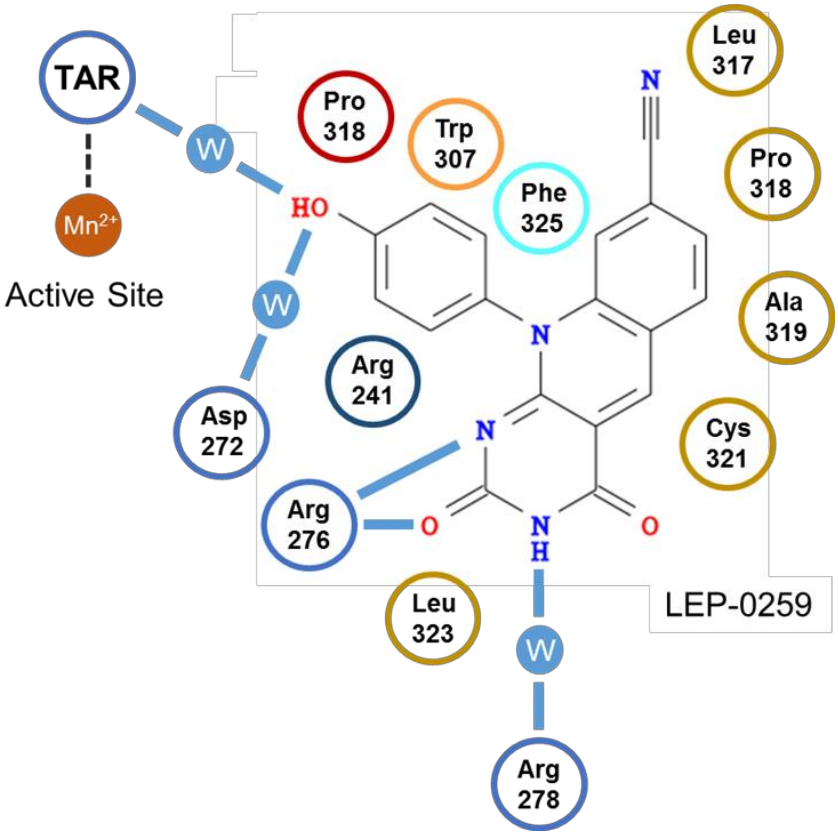

Figure 5.13. LEP-0259 ligand interaction diagram. Key: TAR: tartrate, W: water molecule; blue arrows represent hydrogen bonds; dashed line: coordination bond. Residues involved in hydrogen bond formation and water mediated interaction are circled in blue. Trp308 from the TWDT motif, is circled in orange. Phe325 from the RFDR motif is circled in cyan. Residues circled in yellow indicate residues from the divergent region of the M7-loop. Pro318 (circled in red) is involved in stacking interactions with the compound, but arises from the second protomer in the asymmetric unit.

pointed out to solvent, and does not interact with any residues of the catalytic site. Furthermore, the tetrazole moiety is involved in crystal contacts with a symmetry-related molecule. Moreover, in the m2hTDP2-CAT/C14297 complex, the tetrazole moiety is still pointed out to solvent, but is now no longer involved in any direct crystal contacts. As the two C14297 complexes are found in different crystal forms, we can surmise that the observed position of the tetrazole moiety is not driven by crystallisation contacts, and does not represent a crystallisation artefact; it therefore correctly represents the binding mode of this particular compound.

In contrast, the 3-hydroxyphenol moiety of LEP-0259 points into the active site of TDP2, and makes several hydrogen bonds with key amino acid residues, and also (via a water mediated interaction) interacts with a tartrate molecule trapped at the active site.

In order to improve binding and specificity of the next iteration of compounds, it is likely to be advantageous to extend or 'grow' the molecule towards the active site of TDP2. By adding substituents with functional groups capable of fitting into, or maintaining the extensive hydrogen bond network around the active site, the inhibitory activity of compounds may be increased. A good starting point has been fortuitously provided by the tartrate molecule found bound in the m2hTDP2-CAT/LEP-0259 complex. The position of the tartrate, and the series of water-mediated interactions made with it, provide several 'molecular paths' to explore in future drug design.

In a drug discovery project, small molecule inhibitors are not only tested by biochemical and biophysical assays, but are also examined *in vivo*, to determine whether a compound has the same inhibitory profile in the context of the cell. Cell based assays can provide valuable information about the absorption, bioavailability, metabolism and excretion (ADME) profile, and pharmacodynamics and pharmacokinetic characteristics (PK/PD) of a compound (Meibohm and Derendorf 1997, Doogue and Polasek 2013). These type of assays can aid selection of compounds, from a list of 'hits', which have the desired properties of a 'hit-to-lead' compound.

The next chapter describes some initial experiments looking at whether the deazaflavin series of TDP2 inhibitors function in the context of human cells.

CHAPTER SIX

Cell-based assays with TDP2 inhibitors

6.1 Introduction

In this chapter, preliminary results from two-cell based assays are described: a clonogenic survival assay and a double-strand break repair assay (monitoring γH2AX formation). Both assay formats have already been used in our own laboratory, in order to test the effect of deleting TDP2 in vertebrate cells (Zeng et al. 2011, Zeng et al. 2012). Here, the assays were used to test the effects of both the tetrazole compound C14297 and the hydroxyphenyl derivative LEP-0259 in cells.

6.2 Results

6.2.1 Clonogenic survival of U2OS cells in the presence of C14297

TDP2 is the major, if not only, enzyme capable of resolving TOP2 covalent complexes and protects genome integrity in vertebrate cells (Zeng et al. 2011, Gomez-Herreros et al. 2013). Cells in which TDP2 has been deleted (e.g. avian DT40 and mouse MEFs) have extreme sensitivity to TOP2 poisons, such as etoposide. Consequently, cells treated with etoposide ought to be hypersensitive to inhibitors of TDP2 activity.

The human osteosarcoma cell-line U2OS is commonly employed in cell-based assays, as these cells are large and flat, grow in monolayers, and generate easily countable colonies after 10-14 days in culture. (For the full description Materials and Methods 2.8.3). Cells were pre-treated with C14297 for 4 hr at a final concentration of 1μM, which was far in excess (approximately 20 fold) of the EC50 determined from in vitro experiments, chosen as the maximum dose that would prevent toxicity induced by the DMSO vehicle, and then treated with 0-800nM etoposide for 4 days in continued presence of the inhibitor (Figure 6.1). Control experiments suggested that C14297 is not toxic to cells at 1μM, over the period of the experiment (typically 10-12 days). However, C14297 also had no demonstrable impact on the sensitivity of U2OS cells to etoposide. Similar results were also obtained when experiments were carried out with an increased concentration of C14297 (5μM) (Figure 6.2).

6.2.2 γH2Ax assay in the presence of C14297

To measure the impact of C14297 on DNA double strand break repair (DSBR), we measured the induction and loss of γH2AX immunofoci; an indirect marker of DNA

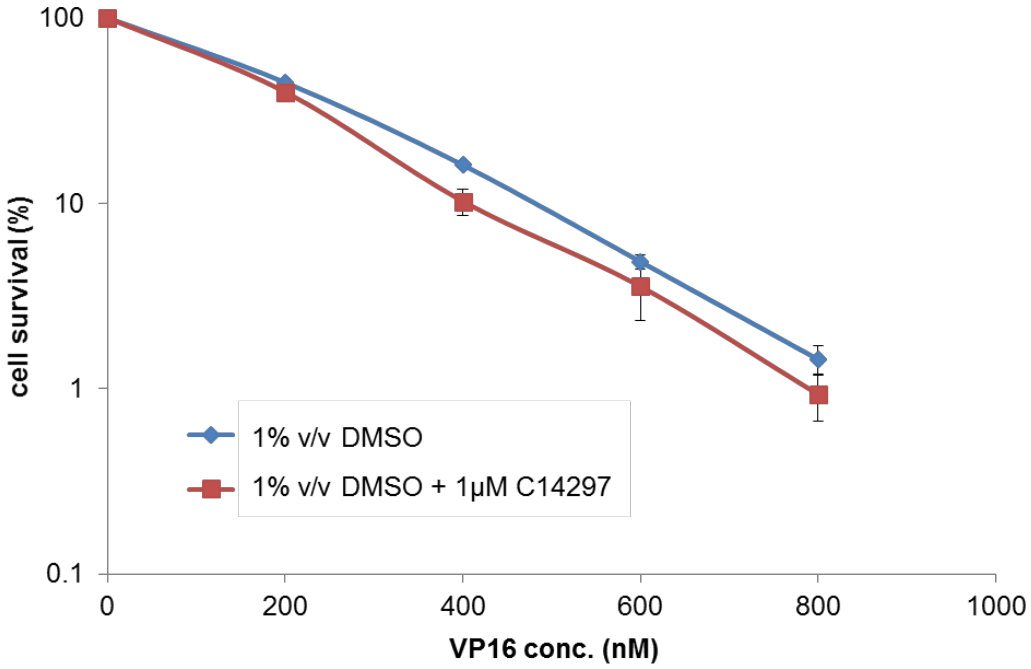

Figure 6.1. Clonogenic survival assays in U2OS cells. Cells were treated with the indicated concentrations of etoposide (VP16) in the presence of DMSO vehicle or 1µM C14297. Data are the mean of three independent experiments, with error bars representing 1 standard deviation. Data points were normalised to the control (100%).

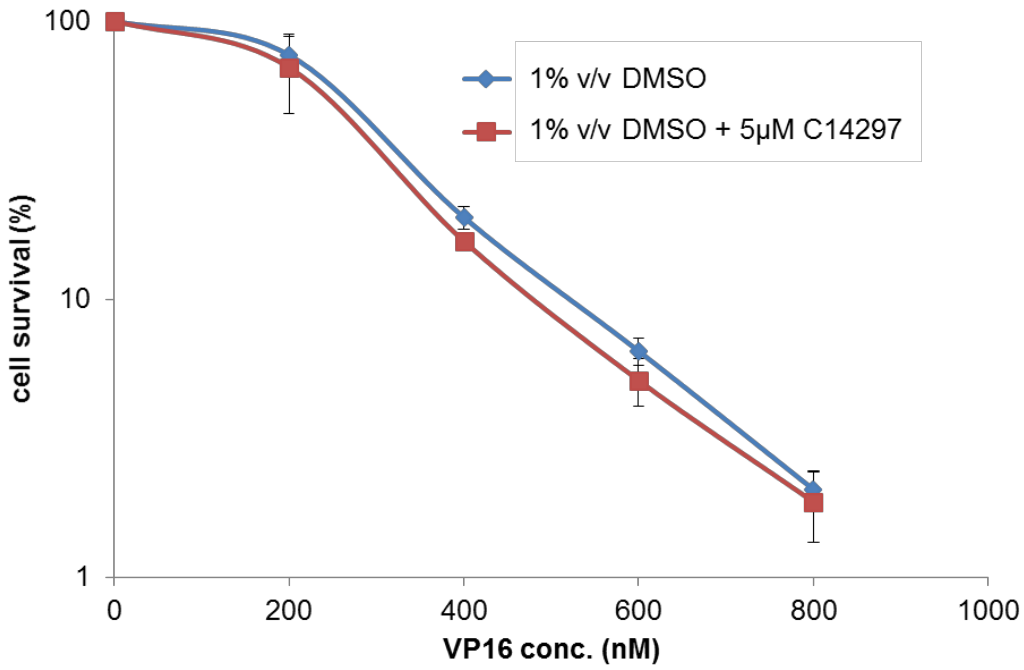

Figure 6.2. Clonogenic survival assays in U2OS cells. Cells were treated with the indicated concentrations of etoposide (VP16) in the presence of DMSO vehicle or 5µM C14297. Data are the mean of three independent experiments, with error bars representing 1 standard deviation. Data points were normalised to the control (100%).

double-strand breaks (DSBs) (Materials and Methods 2.8.4). Human primary fibroblast cells (1BR) were employed for these experiments, because γH2AX foci are readily detected and countable in these cells. 1BR cells were first serum starved to arrest cells in a G1-phase of the cell cycle, thereby omitting S/G2 phase cells from the assay and so enabling the assay to detect specifically NHEJ-mediated DSB repair. Arrested cells were pre-treated with C14297 (5μM) for 1 hour, before the addition of etoposide (20μM) for a further hour. Cells were subsequently then incubated for a further period in etoposide free medium and cells fixed for immune-labelling at the indicated time points (Figure 6.3). However, C14297 did not affect the kinetics of DSBR, under the experimental conditions employed in these experiments.

6.2.3 Assays in the presence of LEP-0259

Similar results were obtained in both assays with LEP-0259. Due to the higher aqueous solubility of LEP-0259, it was possible to carry out the clonogenic survival assay at a higher concentration of compound (10μM). Controls indicated that at this concentration the compound was not toxic to cells. Although the concentration of LEP-0259 tested was much higher, again there was no reduction in colony numbers compared to controls (Figure 6.4). There was no difference between controls and cells treated with LEP-0259 in the γH2Ax focus formation assay (Figure 6.5) at this compound concentration.

6.3 Discussion

Cell based assays are important, especially during the early stages of the drug development process, in order to provide valuable information about the absorption, bioavailability, metabolism, excretion and toxicity (ADME/Tox) profile, and physicochemical and pharmacokinetic (PK) characteristics of a compound (Li 2001). These type of assays can aid selection of compounds, from a list of 'hits', which have the desired properties of a 'hit-to-lead' compound. The tetrazole (C14297) and the hydroxyphenyl derivatives (LEP-0259) of the deazaflavin series of TDP2 inhibitors were tested in clonogenic cell survival and DNA double-strand break repair assays.

In the clonogenic cell survival assays, cells were subjected to continuous treatment with both etoposide and deazaflavin compound. As this assay was previously used to demonstrate that TDP2-depleted and/or deleted cells were extremely sensitive to etoposide treatment (Cortes Ledesma et al. 2009, Zeng et al. 2011), it was reasonable to think that an

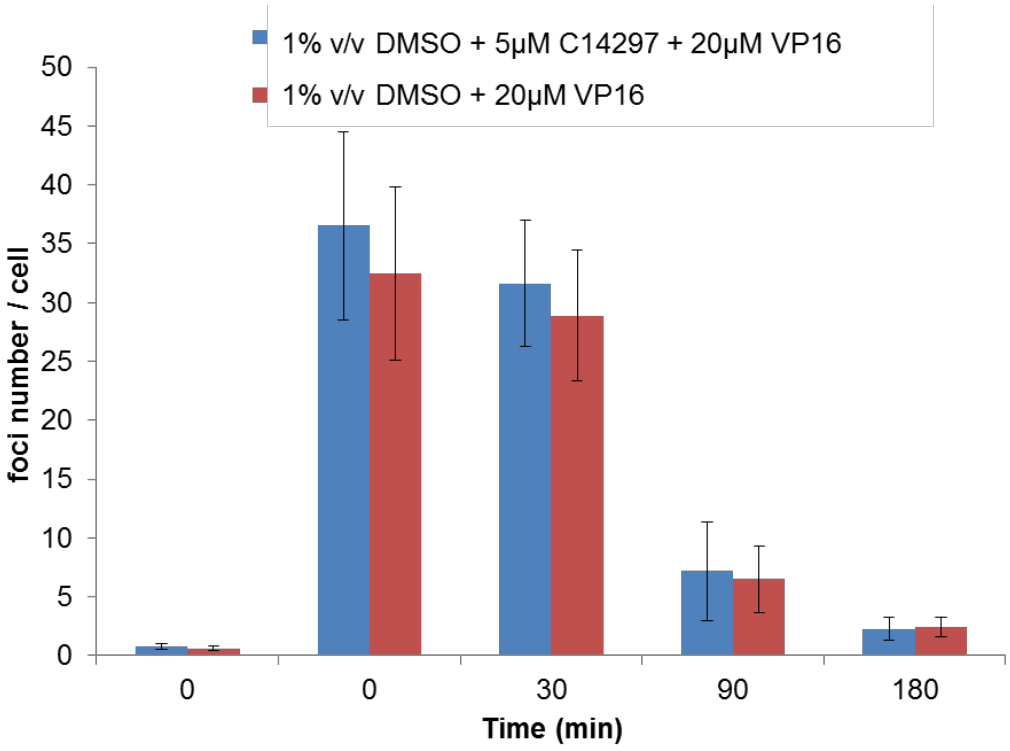

Figure 6.3. γH2AX assays in 1BR cells. Cells were treated with 20μM etoposide (VP16) in the presence of DMSO vehicle or 5μM C14297. Data are the mean of three independent experiments, with error bars representing 1 standard deviation.

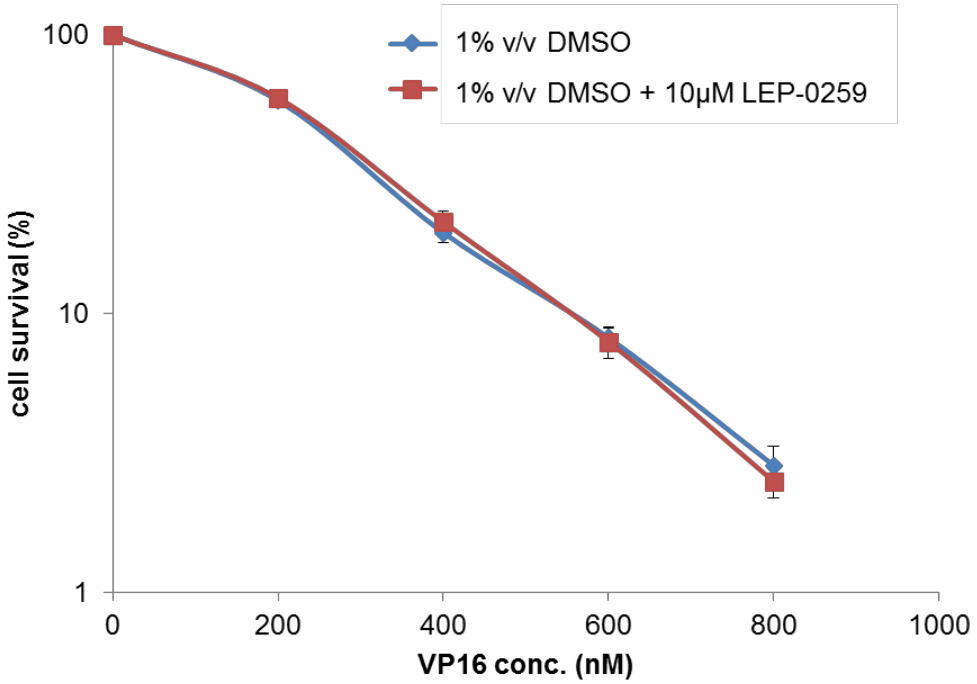

Figure 6.4. Clonogenic survival assays in U2OS cells. Cells were treated with the indicated concentrations of etoposide (VP16) and DMSO vehicle or 10μM LEP-0259. Data are the mean of three independent experiments, with error bars representing 1 standard deviation. Data points were normalised to the control (100%).

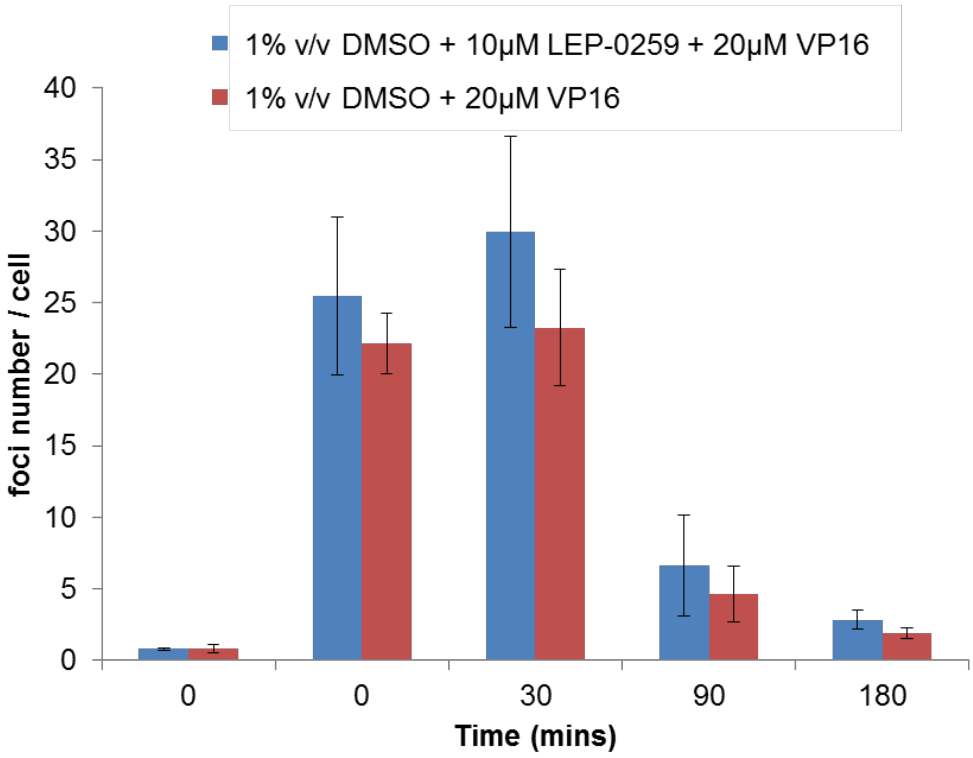

Figure 6.5. γH2AX assays in 1BR cells. Cells were treated with 20µM etoposide (VP16) and DMSO vehicle or 10µM LEP-0259. Data are the mean of three independent experiments, with error bars representing 1 standard deviation.

inhibitor of TDP2 enzymatic activity would produce similar levels of sensitivity. However, neither of the compounds tested resulted in a reduction in cell viability. Similarly, neither compound had any observable effect on the kinetics of DSBR, as measured by the γH2AX focus formation assay.

As the EC50 of each compound is in the micromolar range (C14297: 40nM, LEP-259: 50nM) one would have expected to observe some cellular effects. However, the poor aqueous solubility of the deazaflavin compounds, along with low cell penetrability, may have prevented the compounds from entering the cell. This problem should be addressed directly in future, to find a suitable hit molecule in the future. Several methodologies can be applied to evaluate a drug candidate ability to cross the intestinal barrier and to study their transport mechanism through biological membranes. The Caco-2 cell system has been proven to be an important tool to test drug permeability in drug discovery projects (Angelis and Turco 2011).

CHAPTER SEVEN

The Phosphate-binding pocket of the BRCT1 domain of XRCC1 binds poly(ADP-ribose)

7.1 Introduction

The synthesis of poly(ADP-ribose) (PAR) by PARP1 promotes XRCC1 accumulation at DNA damage sites (Caldecott et al. 1996). It was also hypothesised that the first BRCT domain of XRCC1 (BRCT1) is responsible for this interaction, and binding to PARP1 might be mediated though the poly(ADP-ribose) polymer (PAR). Initial studies attributed the PAR-binding activity to the C-terminal basic region of BRCT1 (Pleschke et al. 2000). This motif consists of two conserved regions: the first rich in basic amino acids and located N-terminally from the second motif comprised of mainly basic and hydrophobic amino acid residues (Figure 1.2.6 A). Similar PAR-binding motifs have been identified in other proteins also involved in DNA repair, such as Ku70, DNA-PKcs and XPA (Pleschke et al. 2000). However, mutations of this motif failed to cause any delay or inhibition of XRCC1 recruitment to DNA strand breaks in cells (unpublished observations, Caldecott lab), suggesting that another motif must be responsible for the interaction.

BRCT domains have a well-established role in phosphorylation-dependent interactions via a highly conserved phosphate-binding pocket (Sheng et al. 2011). Based on examination and comparison of the publically available NMR structure of the BRCT1 domain of XRCC1 and other phosphate binding BRCT domains , it was possible to identify the phosphate-binding pocket, comprised of the conserved amino acid side chains of Ser328, Arg335 and Lys369 (Figure 7.1). Mutation of the latter two amino acids to alanine result in a severe delay in repair and XRCC1 recruitment at DNA damage sites, in cellular assays (Breslin et al. 2015). The aim of the experimental work described in this chapter was to examine whether the interaction between XRCC1 and PARP1 is mediated by the phosphate-binding pocket of the BRCT1 domain.

7.2 Results

7.2.1 XRCC1 co-purifies with PARP1 in a BRCT1 phosphate-binding pocket dependent manner

XRCC1-mutant EM9 cells (which are null for XRCC1) were complemented with expression constructs encoding either full-length C-terminal histidine-tagged wild-type XRCC1 (XRCC1-His), full-length C-terminal histidine-tagged mutant XRCC1 harbouring the phosphate-binding mutations R335A and K369A (XRCC1-His^{RK}), or N-terminal His and Myc tagged truncated XRCC1$^{161-406}$ (His-XRCC1$^{161-406}$) encompassing the BRCT1

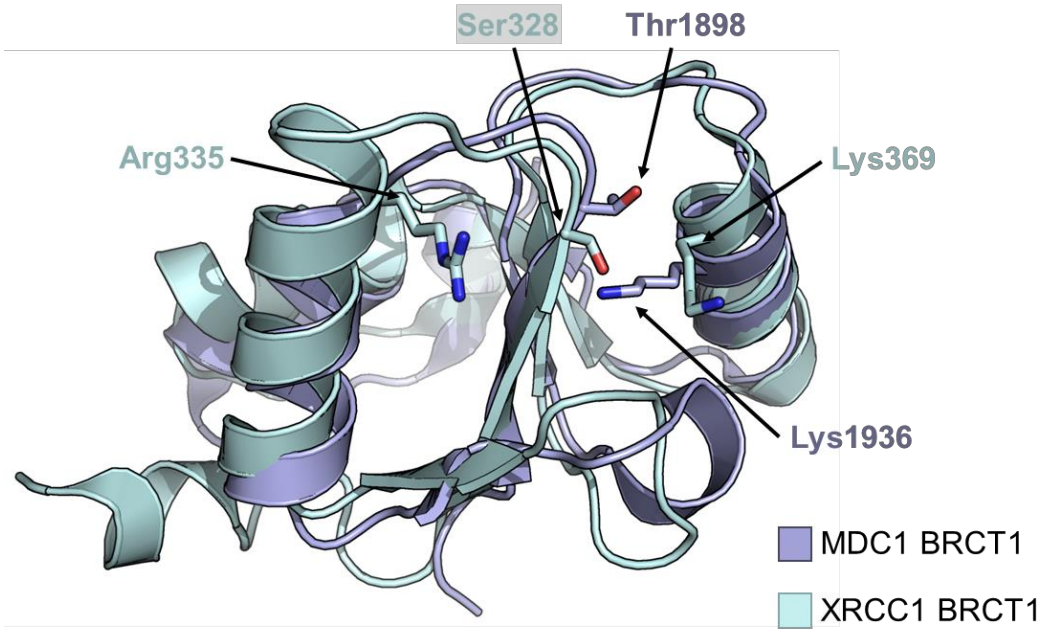

Figure 7.1. Molecular cartoon representation of XRCC1 BRCT1 superimposed on the first BRCT domain of MDC1 (PDB accession code: 2AZM) (Stucki et al. 2005), coloured light blue and pale cyan, respectively. Conserved amino acids involved in phospho-binding interactions are labelled and also shown in stick representation.

domain (Figure 7.2). His-tagged XRCC1 proteins were then recovered from EM9 cell extracts using Ni-agarose resin in affinity 'pull-down' experiments, and probed by western blot for the presence of PARP1 (Materials and Methods 2.7.2) (Figure 7.3).

Both XRCC1-His and His-XRCC1$^{161-406}$ co-precipitated PARP1, suggesting that the extended-BRCT1 domain of XRCC1 is sufficient to bind PARP1 (Figure 7.3). Conversely, the phosphate-binding mutant protein XRCC1-His^{RK} did not pull-down PARP1, supporting the idea that binding is directly mediated by the phosphate-binding motif of BRCT1. Following PARP1 detection, the blot was stripped and re-probed for XRCC1, using either an anti-XRCC1 polyclonal (Bethyl Laboratories) or anti-Myc-tag antibody (Cell Signalling Technology), to detect full-length and truncated XRCC1, respectively (Figure.7.3).

7.2.2 Expression and purification of recombinant His-XRCC1$^{161-406}$, His-XRCC1$^{161-406\ RK}$ and PARP1 from *E. coli*

In the previous section, the interaction between PARP1 and XRCC1 was demonstrated using pull down experiments from human cell lysates. These experiments also highlighted the role of BRCT1 residues R335 and K369 in the interaction, supporting the idea that the interaction is mediated via the phosphate-binding pocket. To examine this interaction in more detail I expressed and purified the recombinant proteins for biochemical assays, in vitro.

DNA encoding amino acid residues 161-406 of XRCC1 was amplified from mammalian expression constructs pCD2E-XH and pCD2E-XH$^{R335A/K369A}$ (Caldecott et al. 1995) and sub-cloned into the bacterial expression vector pTWO-E, using flanking *Nde*I and *Eco*RI restriction sites encoded by the PCR primers (Materials and Methods 2.2.10.1). Successful cloning was confirmed by restriction digest (see Appendix A.3) and *E. coli* strain Rosetta2(DE3) pLysS was employed for subsequent expression of both proteins. The purified proteins are depicted in Figure 7.4A, and a full description of the purification protocol is included in Materials and Methods, (Section 2.2.10.1). To verify that His-XRCC1$^{161-406\ RK}$ was correctly folded, the wild type and mutant recombinant proteins were analysed by fluorescence-based thermal denaturation assays and by circular dichroism spectroscopy (Sections 2.6.1 and 2.6.2). First order derivatives of the thermal denaturation melting curves, or direct fitting of the curves using a modified Boltzmann equation,

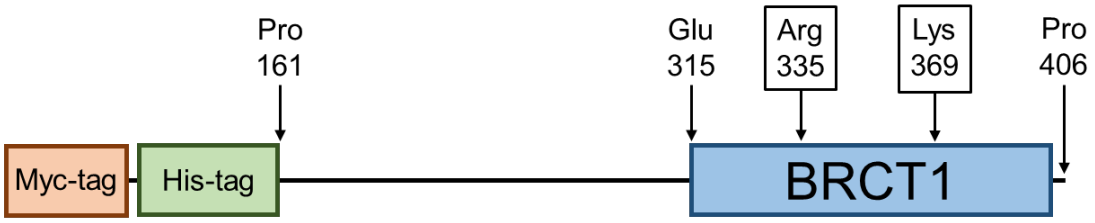

Figure 7.2. Schematic representation of the XRCC1 fragment (spanning residues 161 and 406) used in pull down assays. The constructs harboured an N-terminal Myc- and his-tag for detection and purification purposes. The phosphate-binding mutant harboured the R335A/K369A mutations to abolish phosphate interaction.

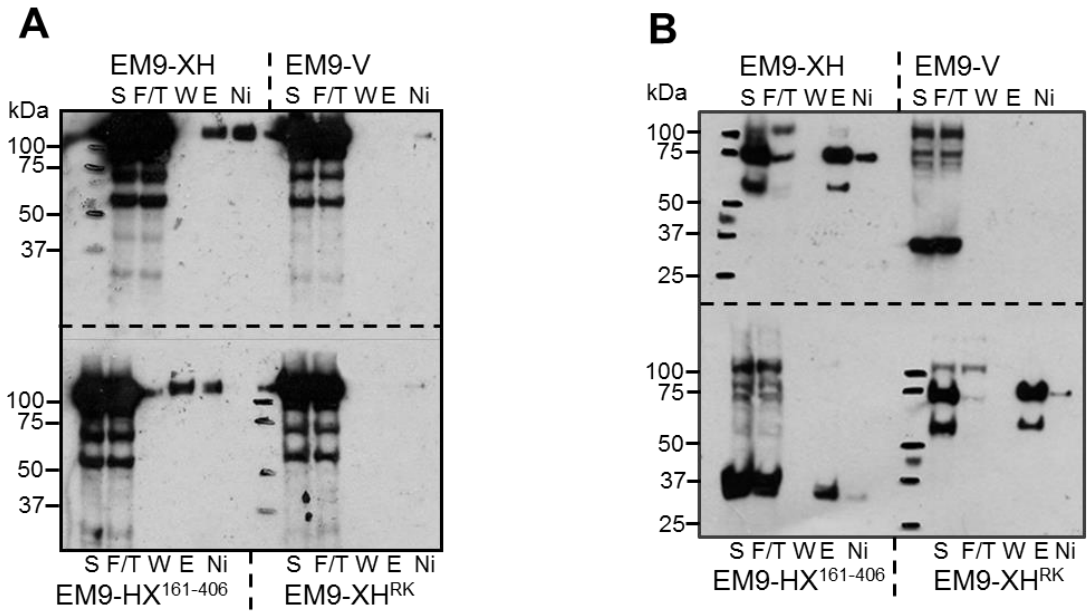

Figure 7.3. The phosphate-binding pocket of the first BRCT domain of XRCC1 binds PARP1. **A.** IMAC pull-down experiments employing cell extracts prepared from XRCC1-mutant EM9 cells stably transfected with expression constructs encoding either full-length wild type human XRCC1-His, XRCC1-His$^{R335A/K369A}$, (EM9-XH^{RK}) or His-XRCC1$^{161-406}$ (EM9-HX$^{161-406}$). Note that the XRCC1 ORFs encode either an N-terminal (truncated XRCC1) or C-terminal (full length XRCC1) His_6-tag. Aliquots of the supernatant (S), flow-through (F/T), final wash (W), the eluent (E), and boiled beads (Ni), were fractionated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted for PARP1 (mouse monoclonal anti-PARP1). **B.** The above membrane was stripped and re-probed for XRCC1, using either the anti-phospho (pS485/pT488) XRCC1 polyclonal antibody or the XRCC1$^{161-406}$ fragment and was detected by mouse c-Myc monoclonal antibody (the Myc tag is N-terminally located on the XRCC1$^{161-406}$ construct).

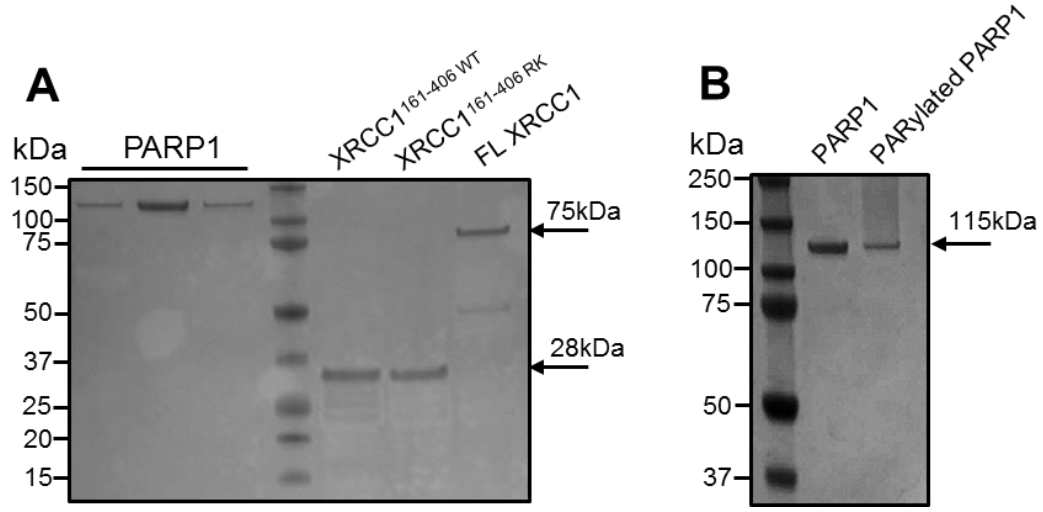

Figure 7.4. Purification of recombinant human His-XRCC1$^{161\text{-}406}$, His-XRCC1$^{161\text{-}406\ RK}$, and PARP1. **A.** PARP1 was expressed in Sf9 cells and purified on Sepharose 4B 3-aminobenzamide resin followed by gel filtration as described in materials and methods. The three peak fractions from the gel filtration step were run on Novex 4-12% Tris-Glycine Protein Gels and then stained with InsantBlue Protein Stain. XRCC1$^{161\text{-}406}$ fragments were expressed in Rosetta^{TM}2 (DE3)pLysS. The initial talon capture was followed by a Heparin column elution. **B.** PARP1 was auto-ribosylated to test enzyme activity. Autoribosylated PARP1 is retarded during electrophoresis resulting in a smear.

indicated that there was no significant difference between the unfolding temperature midpoints of the two proteins, suggesting that the RK-mutant had no significant effect on tertiary structure (Figure 7.5). Similarly, CD spectra of both proteins indicated that the wild type and mutant proteins adopted a similar fold in solution (Figure 7.6).

In addition to recombinant XRCC1, full-length human PARP1 was expressed in, and purified from, Sf9 insect cells as described in Materials and Methods, Section 2.2.5 (Figure 7.4.A). PARP1 activity was verified by incubating the enzyme in the presence of NAD^+ and a single-stranded DNA substrate for 30 min at room temperature, followed by analysis of reaction products on poly-acrylamide gels (Figure 7.4B).

7.2.3 XRCC1 interaction with non-ribosylated and ribosylated PARP1, in vitro.

7.2.3.1 Binding of His-XRCC1$^{161-406}$ and His-XRCC1$^{161-406\ RK}$ to PARP1 on slot blots.

To examine if XRCC1 interacted with PARP1 *in vitro*, I first employed slot blot methodology. Recombinant His-XRCC1$^{161-406\ WT}$ and His-XRCC1$^{161-406\ RK}$ were adsorbed onto nitrocellulose membrane and then incubated with either mock-ribosylated or auto-ribosylated PARP1 (Materials and Methods 2.5.1.1) (Figure 7.7). This experiment strongly suggested that both the wild type and mutant XRCC1 proteins bound auto-ribosylated PARP1 with higher affinity than non-ribosylated PARP. However, these experiments did not support the idea that PAR binding required the phosphate-binding pocket, because wild type and mutant XRCC1 appeared to bind PARP1 with similar efficiency.

7.2.3.2 Binding of His-XRCC1$^{161-406}$ and His-XRCC1$^{161-406\ RK}$ to PAR on slot blots

In the next set of experiments, binding of His-XRCC1$^{161-406}$ to naked PAR was investigated (Materials and Methods 2.5.1.2). A dilution series of PAR polymer (Trevigen, Abingdon, UK) was prepared in PBS and blotted onto a nylon membrane (GE Healthcare). After UV-crosslinking and blocking in non-fat dried milk, the membrane was incubated with His-XRCC1$^{161-406\ WT}$ or His-XRCC1$^{161-406\ RK}$ (Figure 7.8). In contrast to the previous result, this experiment supported the idea that PAR binding is mediated by the phosphate-binding pocket. To resolve this discrepancy, I established a more and quantitative and reproducible assay, using micotitre plate-binding assays that lend themselves to ELISA analysis.

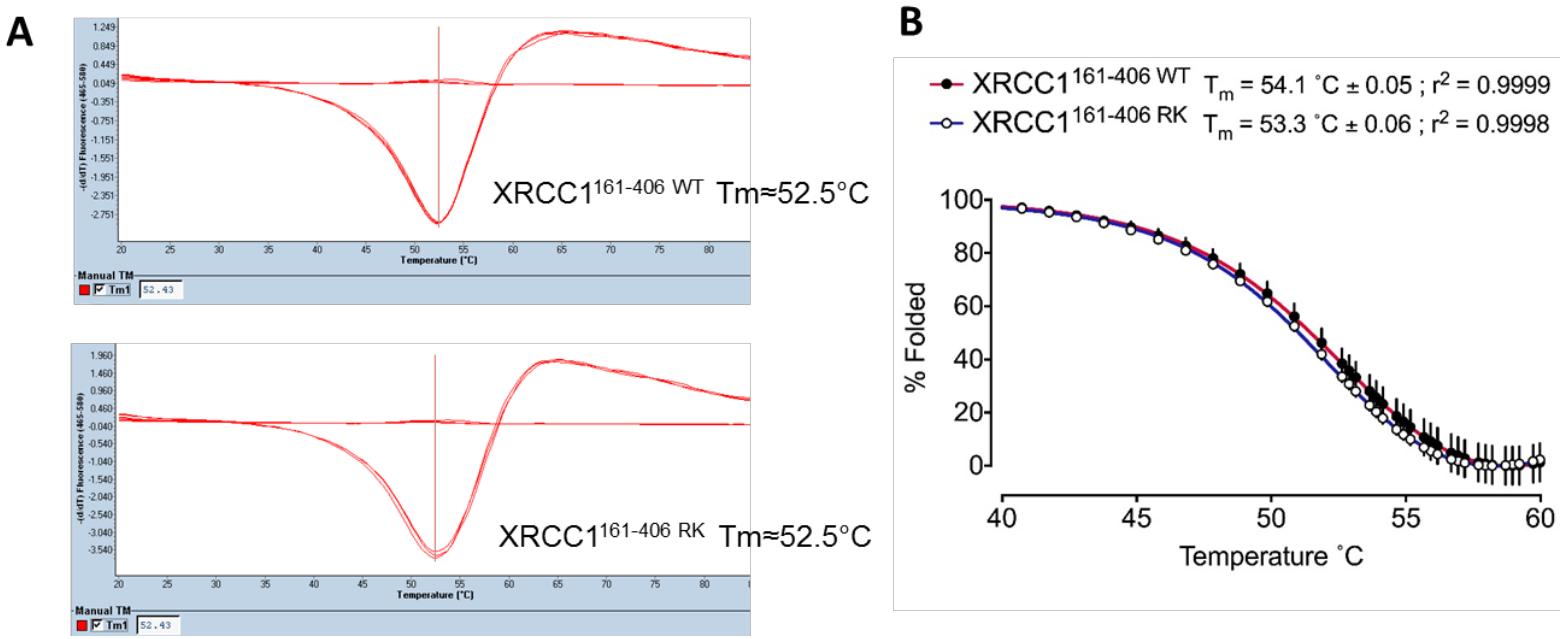

Figure 7.5. Thermal denaturation assay to assist the stability of the XRCC1$^{161\text{-}406}$ fragments. **A.** Unfolding temperatures (Tm) were established by either the first order derivatives of the thermal denaturation melting curves, or, **B.** Direct fitting of the curves using a modified Boltzmann equation.

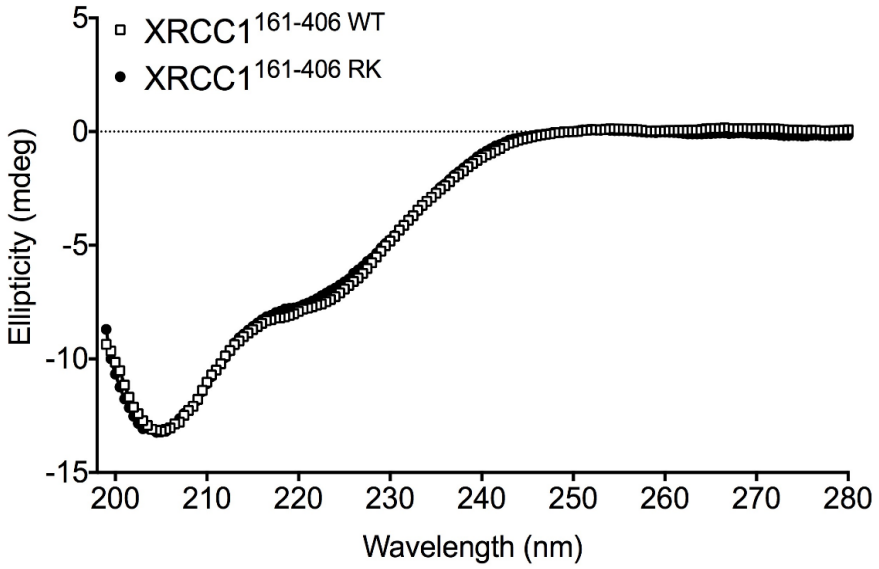

Figure 7.6. CD spectra for XRCC1$^{161\text{-}406\ WT}$ and XRCC1$^{161\text{-}406\ RK}$. Protein samples in 10mM HEPES, pH7.5, 300mM NaCl and 0.5mM TCEP were concentrated to 1.55mg/ml. Spectra were measured between the wavelengths 198nm and 280nm.

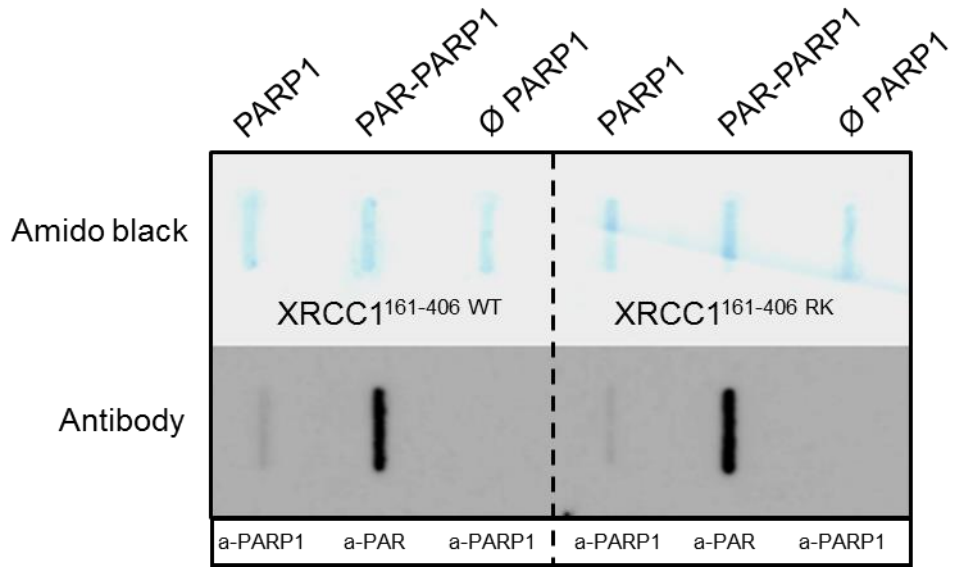

Figure 7.7. XRCC1 BRCT1 domain binds PARylated PARP1. 15µg of either wild type His-XRCC1$^{161-406}$ or His-XRCC1$^{161-406\ RK}$ were absorbed onto nitrocellulose membrane. The membrane was incubated in the presence of either 100nM ribosylated, or non-ribosylated PARP1. Non-ribosylated PARP1 was detected by mouse anti-PARP1 antibody (a-PARP1) (Bio-Rad), and ribosylated PARP1 ("PAR-PARP1") was detected by mouse anti-PAR antibody (a-PAR) (Enzo Life Sciences). Amido black staining shows equal loading of XRCC1$^{161-406}$ proteins.

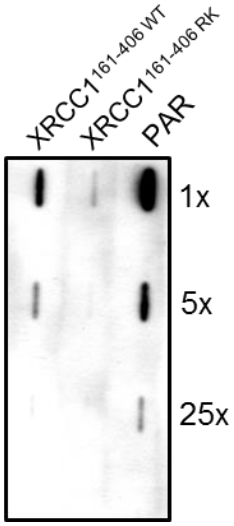

Figure 7.8. XRCC1 binding experiment to naked PAR polymer. Dilution series of naked PAR polymer (10, 2, and 0.4ng) were absorbed onto Hybond-XL nylon membrane (GE Healthcare) and were incubated with either WT or the R335A/K369A mutant XRCC1$^{161-406}$ fragment (100nM). Proteins were detected by anti-polyhistidine antibody. In the last lane, poly(ADP-ribose) was detected by anti-poly(ADP-ribose).

7.2.3.3 Binding of His-XRCC1$^{161\text{-}406\ WT}$ and His-XRCC1$^{161\text{-}406\ RK}$ to ribosylated proteins in a plate-binding assay

A colorimetric-based assay was established previously in the laboratory to measure protein binding to ribosylated proteins (Rulten at al., 2008), which I adapted here for my experiments (Figure 7.9). A full description of the protocol is included in the Materials and Methods (section 2.5.2.1). However, in brief, a 96 well plate is coated with calf thymus histone type II (Sigma-Aldrich), PARP1, or other proteins of interest (e.g. BSA), and the histones or control proteins then either mock-ribosylated or trans-ribosylated by recombinant PARP1. After extensive washing to remove unbound proteins, the ability of His-XRCC1$^{161\text{-}406}$ to bind the mock-ribosylated or ribosylated histones or control proteins was measured by incubating the recombinant His-XRCC1$^{161\text{-}406}$ with the coated well for 30 min. The wells were then extensively washed and bound His-XRCC1$^{161\text{-}406}$ detected using a mouse monoclonal anti-polyhistidine antibody (Sigma Aldrich), anti-mouse alkaline-phosphatase conjugated secondary (DAKO), and TACS Sapphire detection reagent (Trevigen).

In this assay, His-XRCC1$^{161\text{-}406\ RK}$ was less able to bind auto-ribosylated PARP-1, trans-ribosylated calf thymus histone H1, or trans-ribosylated Bovine Serum Albumin (BSA) than was His-XRCC1$^{161\text{-}406}$. This was most evident when low concentrations of NAD^+ were employed to limit the total level of PAR synthesis (Figure 7.10). These data strongly support the notion that the phosphate-binding pocket of the BRCT1 domain is required for PAR binding by XRCC1, at low levels of PAR synthesis at least. This in turn supports a model in which the phosphate-binding pocket of the BRCT1 domain binds the phosphate moieties present in the PAR polymer directly (Figure 7.1B). One potential caveat is that the recombinant PARP1 and histone employed here derived from insect cells and calf thymus respectively, raising the possibility that these proteins were phosphorylated and that the phosphate-binding pocket found in XRCC1 bound this phosphate, albeit in a ribosylation dependent manner. To rule out this possibility, the experiment was repeated using recombinant histone H1 expressed in *E. coli*, with similar results (Figure 7.11).

7.2.3.4 XRCC1$^{161\text{-}406}$ binds poly(ADP-ribose) but not mono(ADP-ribose)

The plate-binding assays suggested that the BRCT1 phosphate-binding pocket interacts directly with PAR. A major question that arose in light of this result was what is the minimal chain length required for efficient BRCT1 binding? The basic unit of poly (ADP-ribose) is adenosine diphosphate ribose (ADP-ribose; ADPr). It was hypothesised that if

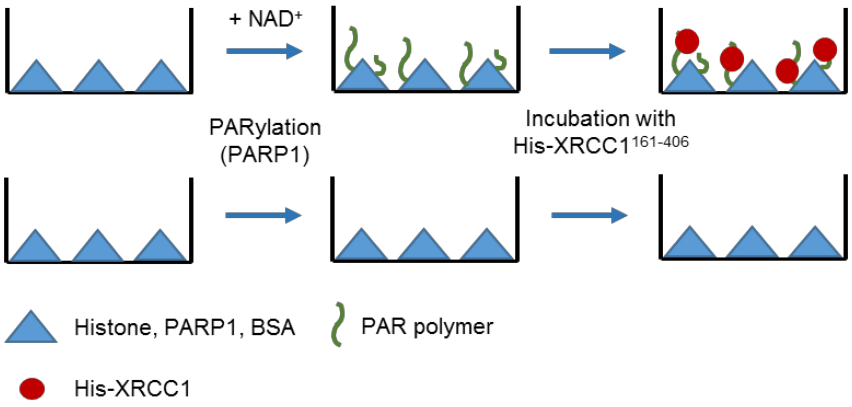

Figure 7.9. Schematic representation of the plate-binding assay. First, histones, PARP1, or BSA (as a control) were absorbed onto the wells of a 96 well plate. After extensive washing, the proteins were either mock-ribosylated or ribosylated in the presence of various concentration of NAD^+, rinsed extensively and incubated with wild type or mutant His-$XRCC1^{161-406}$ for 30 min. After washing, His-XRCC1 detected by polyhistidine antibody as described in materials and methods (2.5.2.1).

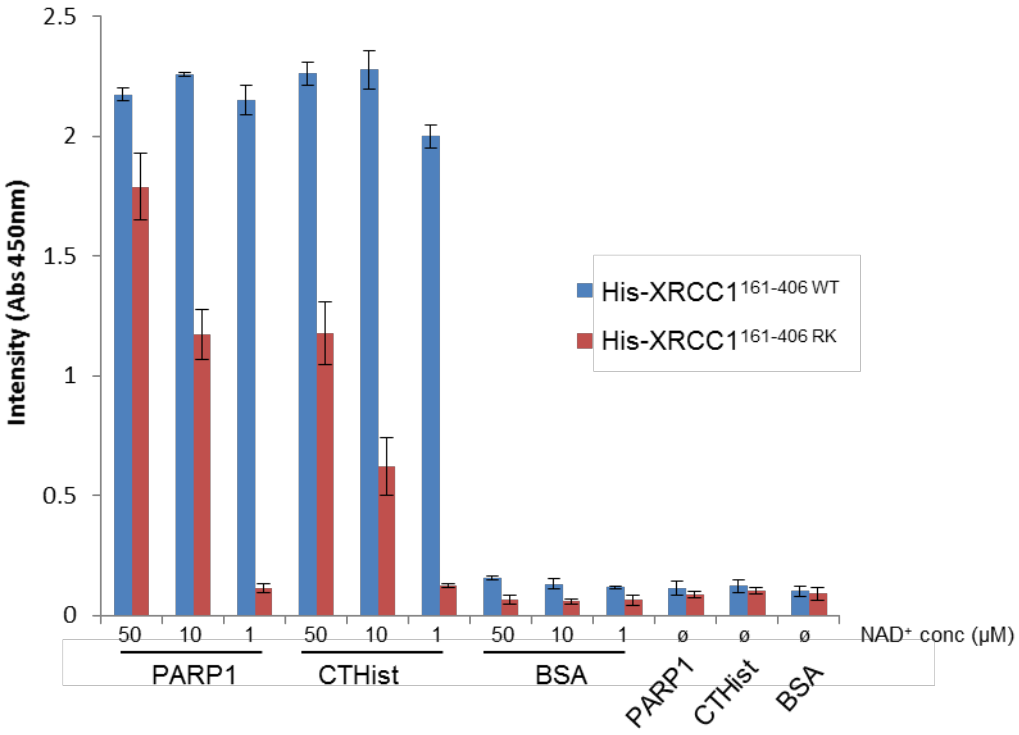

Figure 7.10. The XRCC1 BRCT1 phosphate-binding pocket binds PAR, in vitro. Binding of His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$ to the indicated mock-ribosylated (-NAD$^+$) or ribosylated (1–50μM NAD$^+$) proteins was measured as indicated in Figure 7.9. Data are the mean (±1 SD) of at least three experiments. KEY: CTHist: calf thymus histone, BSA: bovine serum albumin.

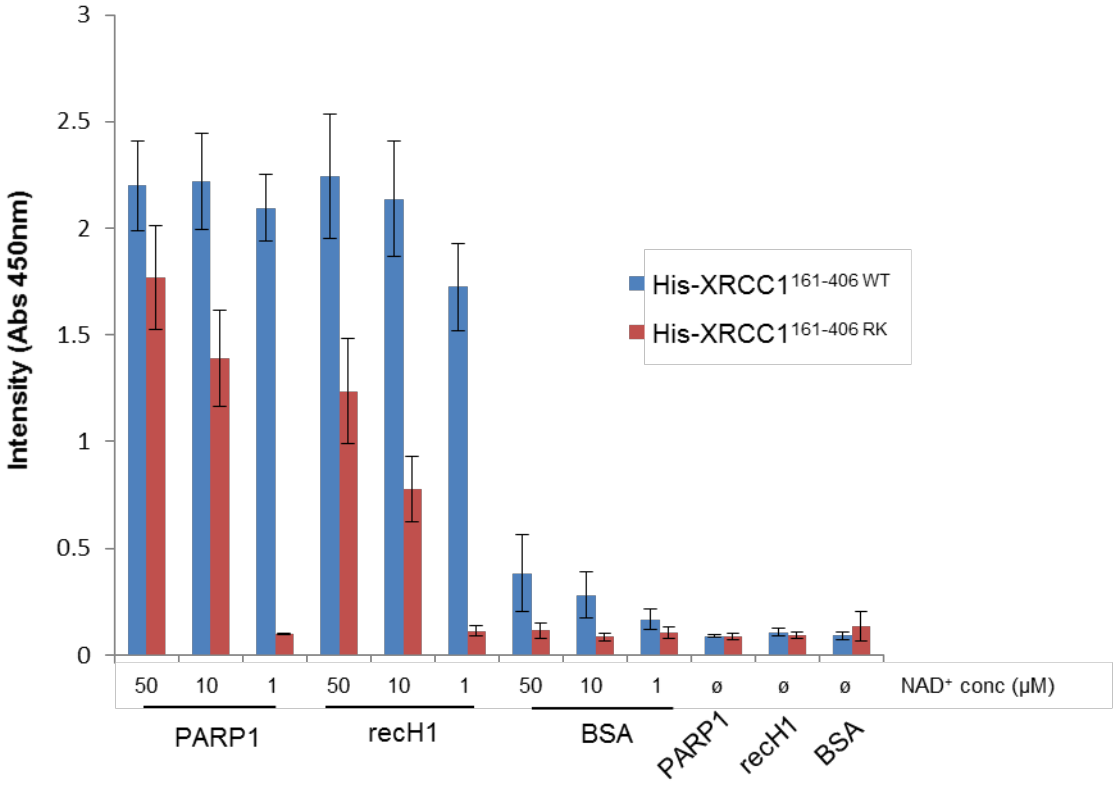

Figure 7.11. The XRCC1 BRCT1 phosphate-binding pocket binds PAR, in vitro. Binding of His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$ to the indicated mock-ribosylated (-NAD$^+$) or ribosylated (1–50μM NAD$^+$) proteins was measured as indicated in Figure 7.9. Data are the mean (±1 SD) of at least three experiments. KEY: recH1: recombinant histone H1, BSA: bovine serum albumin.

the phosphate-binding pocket binds a single ADPr unit (mono ADPr) then it should be outcompeted by an increasing concentration of free unbound ADPr. To test this, different concentrations of mono ADP-ribose competitor (50nM, 0.5μM, 50μM, 500μM) were pre-incubated with 25nM His-XRCC1$^{161-406}$ before incubation with the adsorbed ribosylated histones (Materials and Methods 2.5.2.3). After extensive washing, the bound His-XRCC1$^{161-406}$ was detected as previously described (section 2.5.2.1). Notably, mono(ADP-ribose) (MAR) failed to prevent His-XRCC1$^{161-406}$ binding, even when present at 1000-fold molar excess (Figure 7.12). I conclude from these experiments that phosphate-binding pocket binds PAR, but not mono(ADP-ribose). To further test this hypothesis, I employed naked poly(ADP-ribose) (PAR; Trevigen), comprised of heterogeneous polymers of 2 to 300 units in length, as a competitor, along with auto-ribosylated PARP1 (Materials and Methods 2.5.2.3). His-XRCC1$^{161-406}$ was pre-incubated with different competitor concentrations prior to adding to the adsorbed ribosylated histones, as above. Unlike MAR, both PAR and auto-ribosylated PARP1 competed effectively with His-XRCC1$^{161-406}$ for binding to ribosylated histone protein (Figure 7.13).

7.2.3.5 The use of biotinylated NAD$^+$ to measure PAR synthesis

To further examine the necessity for PAR, versus MAR, for XRCC1 binding, I developed an assays using biotin-NAD$^+$ to label the ADP-ribose products of PARP activity, directly (Materials and Methods 2.5.2.4). This was necessary to measure the polymer product of PARP enzymes, since available antibodies fail to detect MAR, or short chains of PAR. To test His-XRCC1$^{161-406}$ binding to biotinylated-PAR, calf thymus histone type II (Sigma-Aldrich) was either mock-ribosylated or ribosylated in the presence of biotinylated-NAD$^+$ (1, 10, or 50μM) with PARP1. The experiment was conducted in a duplicate to detect both biotin-NAD+ (using streptavidin) and His-XRCC1$^{161-406}$ binding (using anti-polyhistidine antibody). His-XRCC1$^{161-406}$ also bound the biotinylated substrate, and the level of biotin incorporation was dependent on the NAD$^+$ concentration, as expected (Figure 7.14). Importantly, the binding by XRCC1 was again dependent on the phosphate-binding pocket (Figure 7.15). The overall level of XRCC1 binding was less than observed in experiments employing radiolabelled NAD$^+$ however, as suggested by the total absorbance values. This most likely reflects that biotinylated-NAD$^+$ might be not as efficient substrate for PAR synthesis. Alternatively, the biotin moiety might interfere with the XRCC1 interaction due to steric hindrance. It is important to point out that the signal emerging from the biotin

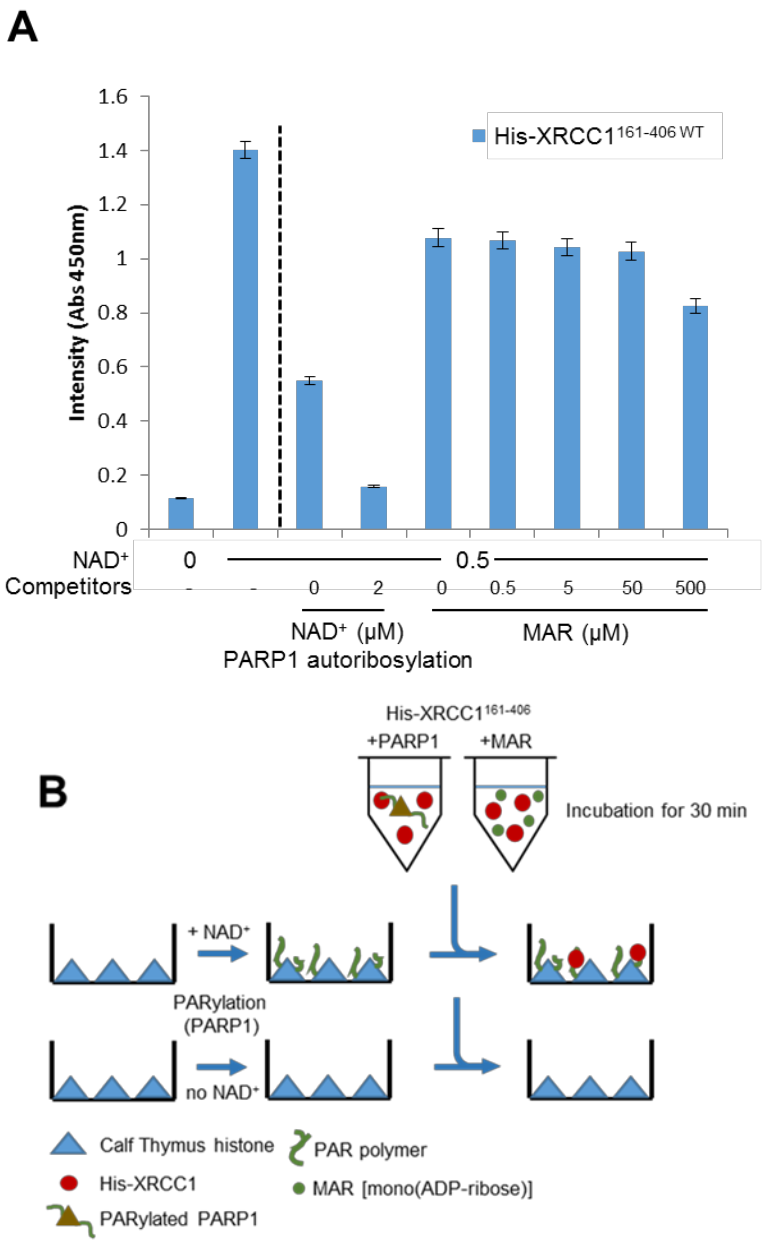

Figure 7.12. The XRCC1 BRCT1 phosphate-binding pocket does not bind mono(ADPr-ribose) (MAR). **A.** Binding of His-XRCC1$^{161–406}$ to calf thymus histone mock-ribosylated in the absence of NAD$^+$ ('0') or ribosylated in the presence of 0.5 μM NAD$^+$. Where indicated, XRCC1 binding was measured in the presence of 43 nM PARP1 competitor that was first autoribosylated in the presence of 0 or 2μM NAD$^+$, as indicated. Alternatively, His-XRCC1$^{161–406\ WT}$ binding was measured the presence of the indicated concentration of mono (ADP-ribose) ('MAR') competitor. MAR competitor concentrations are total ADP-ribose units (μM) present as MAR. Data are the mean (±1SD) of at least three experiments. **B.** Schematic representation of the experiment.

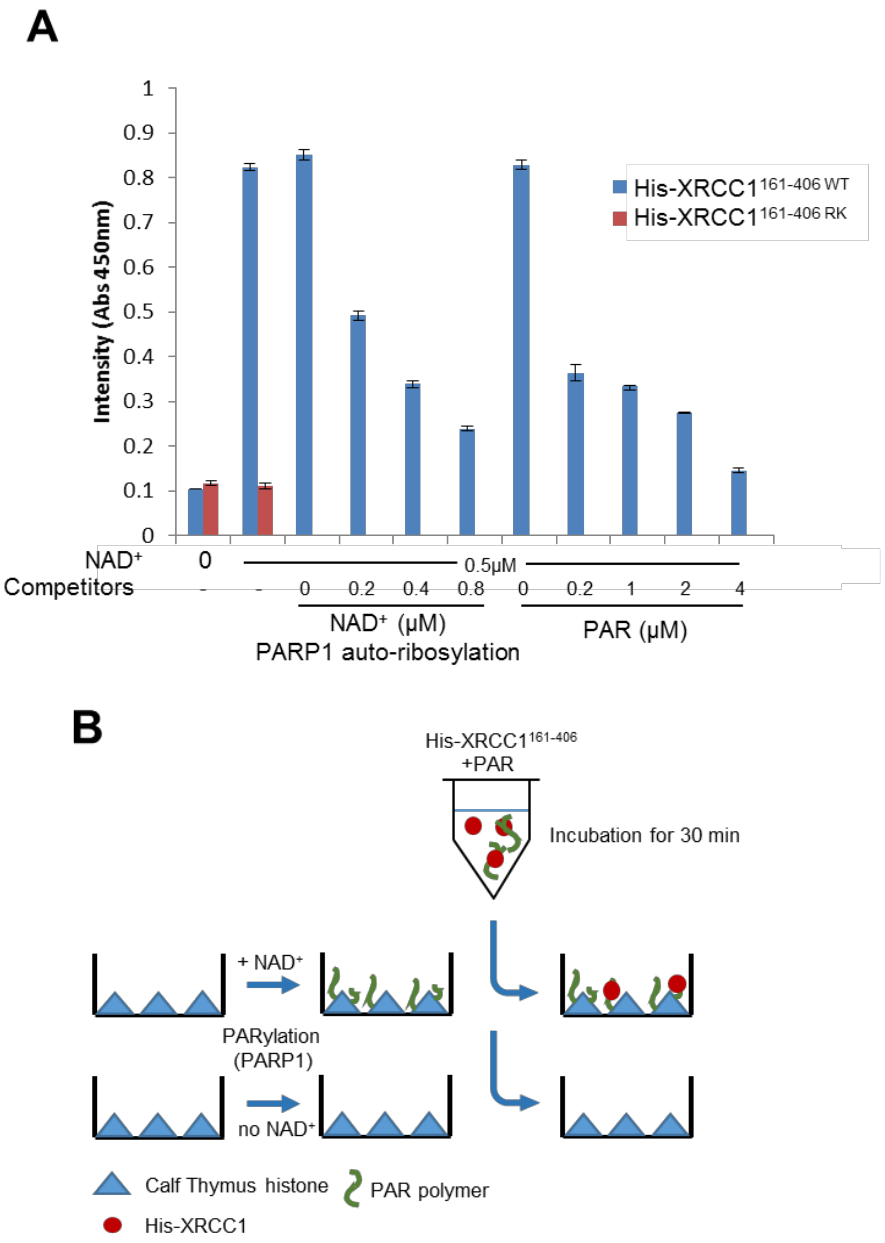

Figure 7.13. The XRCC1 BRCT1 phosphate-binding pocket bind PAR **A.** Binding of either His-XRCC1$^{161–406\ WT}$ or His-XRCC1$^{161–406\ RK}$ to calf thymus histone mock-ribosylated in the absence of NAD^+ ('0') or ribosylated in the presence of 0.5 µM NAD^+. Where indicated, XRCC1 binding was measured in the presence of 43nM PARP1 competitor that was first auto-ribosylated in the presence of 0, 0.2, 0.4 and 0.8µM NAD^+, as indicated. Alternatively, His-XRCC1$^{161–406\ WT}$ binding was measured in the presence of the indicated concentration of PAR competitor. PAR competitor concentrations are total ADP-ribose units (µM) present as PAR (2-300 subunit lengths). Data are the mean (±1SD) of at least three experiments. **B.** Schematic representation of the experiment.

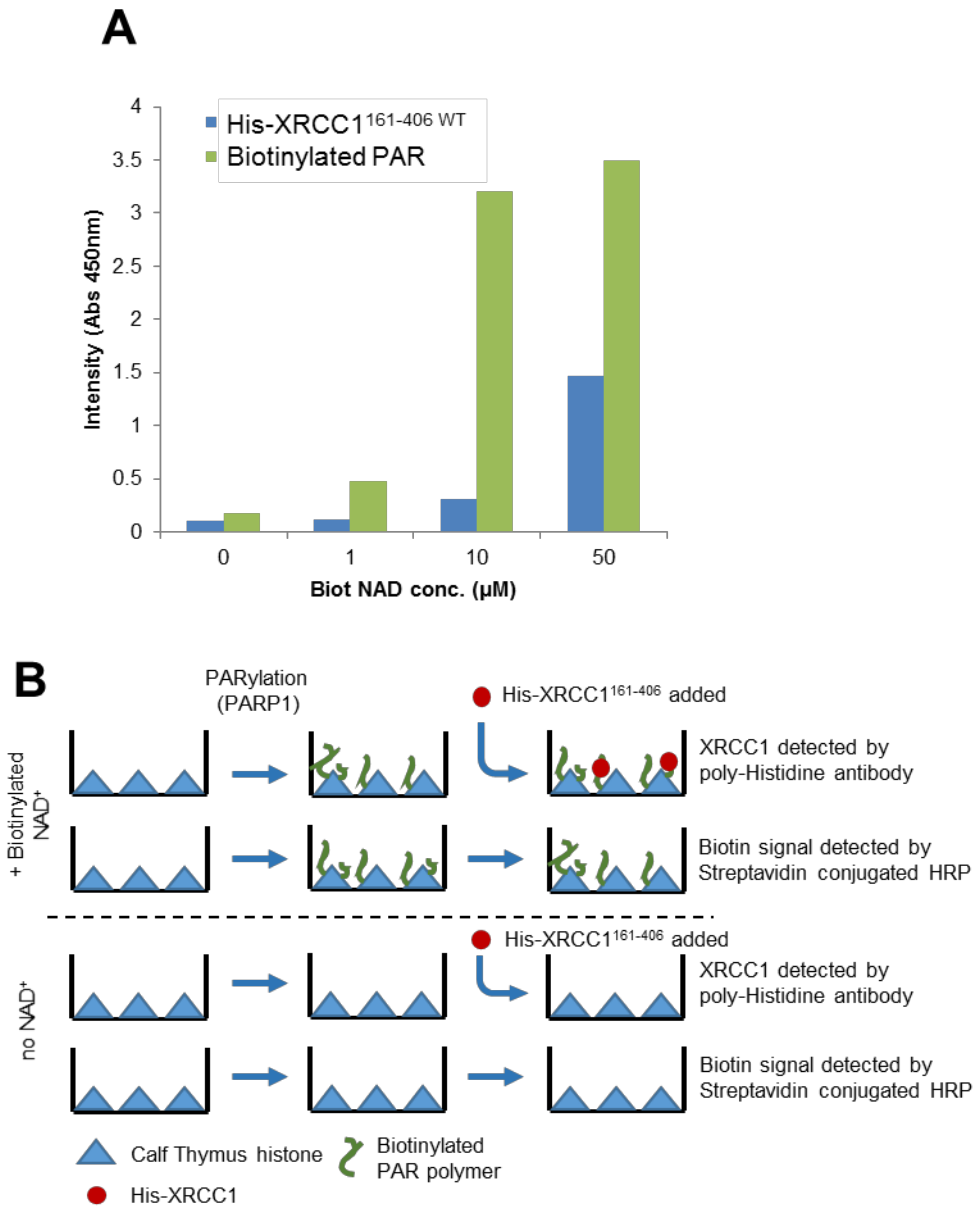

Figure 7.14. The phosphate-binding pocket of the first BRCT domain of XRCC1 binds biotinylated PAR polymer. Calf thymus histone was either mock ribosylated or ribosylated with biotinylated NAD^+ in the presence of PARP1. 25nM wild type $XRCC1^{161-406}$ was added to the wells and incubated for 30 min. The binding of XRCC1 fragment was detected as described in Materials and methods. Additionally, in a parallel experiment the synthetized biotinylated PAR polymer was detected using HRP conjugated Streptavidin. **B.** Schematic representation of the experiment, showing parallel detection of the signals emerging from the biotinylated ADPr moieties incorporated into the forming PAR polymer and the bound $XRCC1^{161-406}$.

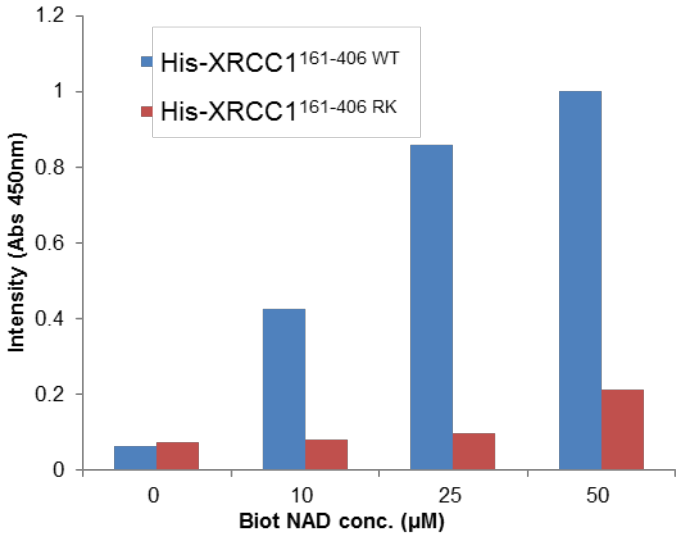

Figure 7.15. Biotinylated PAR binds XRCC1$^{161\text{-}406\ WT}$ and XRCC1$^{161\text{-}406\ RK}$. Calf Thymus histone was ribosylated in the presence of biotinylated NAD^+ with PARP1. 25nM XRCC1 fragment was added and incubated for 30 min on ice, then detected as described in Materials and Methods (2.5.2.1).

moiety detected by HRP-conjugated streptavidin antibody saturates the assay already at 10μM substrate concentration.

7.2.3.5.1 Binding of His-XRCC1$^{161-406}$ to products of PARP3 activity

In the previous section I demonstrated that His-XRCC1$^{161-406}$ can bind to biotinylated PAR synthesised by PARP1. Next, I employed PARP3, which primarily mono-ribosylates proteins, to ribosylate calf thymus histone type II (Sigma-Aldrich), in the presence of 10-100μM biotinylated NAD^+ (Materials and Methods 2.5.2.4). As expected for an enzyme that primarily mono-ribosylates proteins, His-XRCC1$^{161-406}$ was unable to bind the product of PARP3 activity, despite clear biotin-NAD^+ dependent ribosylation of calf thymus histone proteins (Figure 7.16).

7.2.3.5.1 Ribosylated histones do not recruit XRCC1 after PARG treatment

To prove that XRCC1 requires more than one ADPr unit for binding, the biotinylated products of histone ribosylation by PARP1 were pre-treated with enzymes that degraded either PAR (e.g. PARG) or MAR (e.g. MACROD1; MacD1) prior to incubation with His-XRCC1$^{161-406}$ (Materials and Methods 2.5.2.4). Notably, binding of His-XRCC1$^{161-406}$ was prevented by PARG treatment at all NAD^+ concentrations tested (Figure 7.17). In contrast, MACROD1 treatment did not affect His-XRCC1$^{161-406}$ binding to PARylated histones (Figures 7.18). Biotin detection confirmed that PARG greatly reduced biotin label, but did not ablate it completely, as expected given the residual protein-proximal MAR left behind by PARG (Figure 7.19). Together, these experiments confirm that the phosphate-binding pocket of XRCC1 binds PAR, but not MAR.

7.2.3.4 Impact of PAR length on His-XRCC1$^{161-406}$ binding

In order to pursue a structural biology approach to examine the molecular nature of the interaction between the His-XRCC1$^{161-406}$ and the PAR polymer, distinct unit lengths of PAR were synthetized and purified (Material and Methods 2.5.3.2).

Because the chemical synthesis of homogenous PAR polymer with distinct unit length is still challenging, the only viable approach was to obtain PAR from enzymatic reactions employing a poly(ADP-ribose) polymerase. Tan et al presented a protocol to scale up the preparation of homogenous PAR polymer (Tan et al. 2012), which I followed with minor modifications (See Material and Methods 2.5.3). This approach resulted in recovery of numerous distinct UV absorbance peaks during chromatography (Figure 7.20). Peak 1-9,

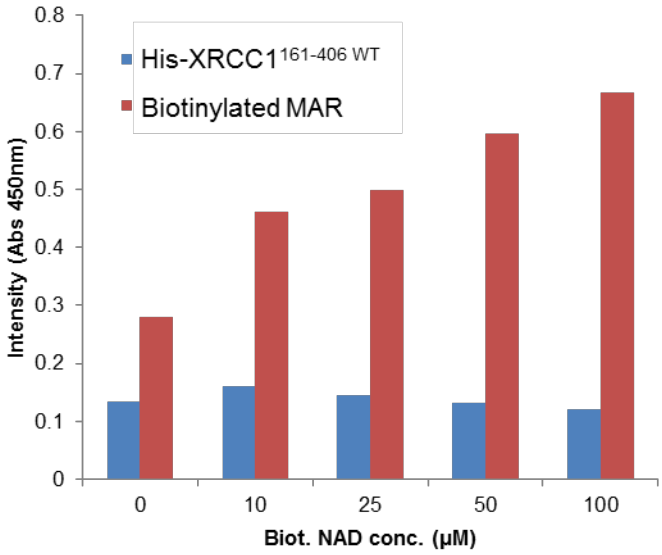

Figure 7.16. The phosphate-binding pocket does not bind mono(ADP-ribose). Calf thymus histone was either mock ribosylated or ribosylated with biotinylated NAD+ in the presence of PARP3. 25nM wild type XRCC1$^{161-406}$ was added to the wells and incubated for 30 min. The binding of XRCC1 fragment was detected as described in Materials and methods (2.5.2.1). Additionally, in a parallel experiment the synthetized biotinylated mono(ADP-ribose) moiety was detected using HRP conjugated Streptavidin.

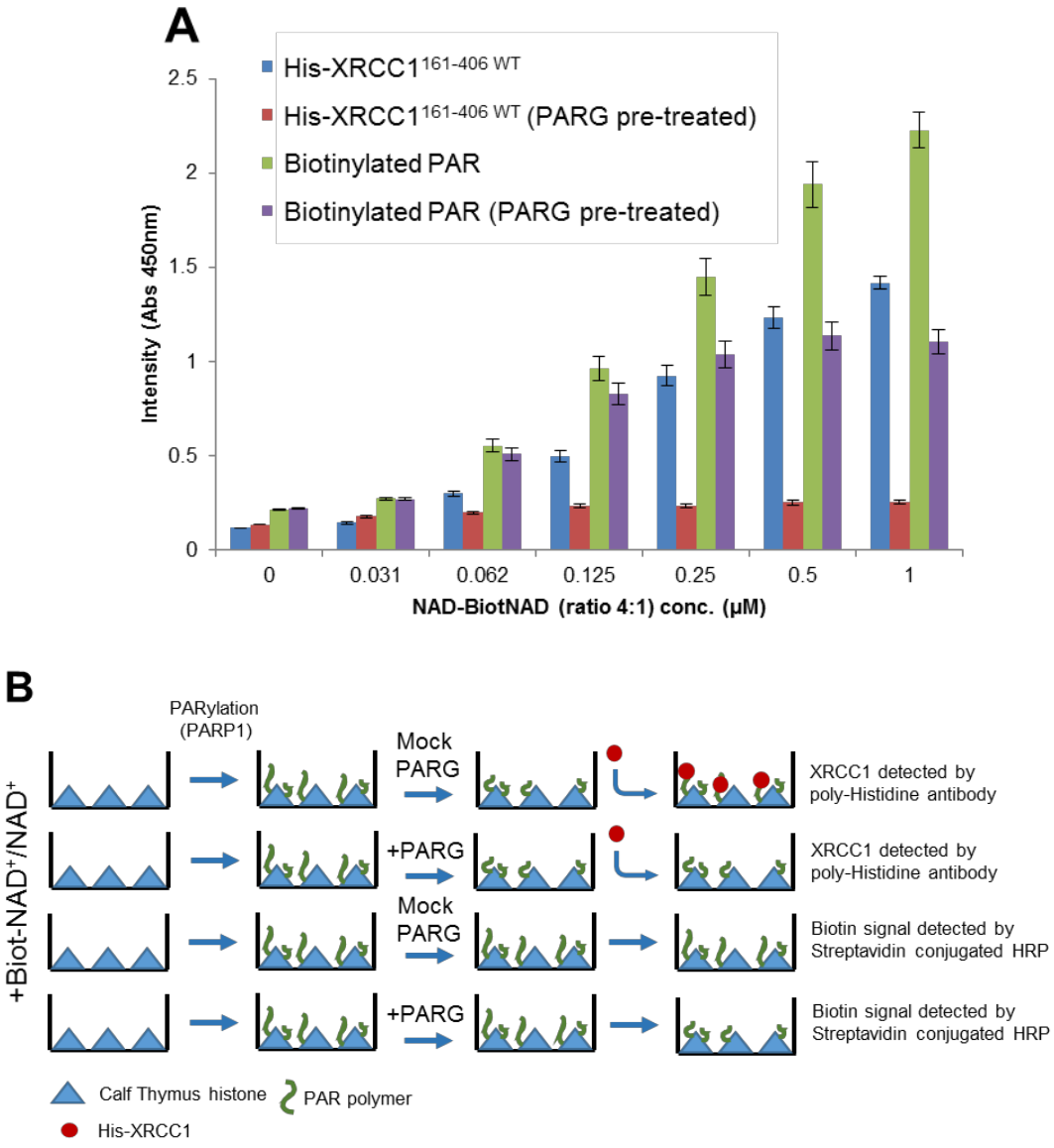

Figure 7.17. The phosphate-binding pocket of the first BRCT domain of XRCC1 binds PAR polymer and not MAR. **A.** For substrate, a mix of NAD^+ and biotinylated NAD^+ in a molar ratio of four to one was employed. The ribosylated proteins were either mock-treated or treated with PARG, then incubated with $XRCC1^{161\text{-}406\ WT}$. The binding of $XRCC1^{161\text{-}406\ WT}$ was detected as described in Materials and Methods (2.5.2.1). In a parallel experiment the level of ribosylation before and after PARG treatment was also detected using HRP-conjugated streptavidin. Data are the mean (±1SD) of three independent experiments. **B.** Schematic representation of the experiment.

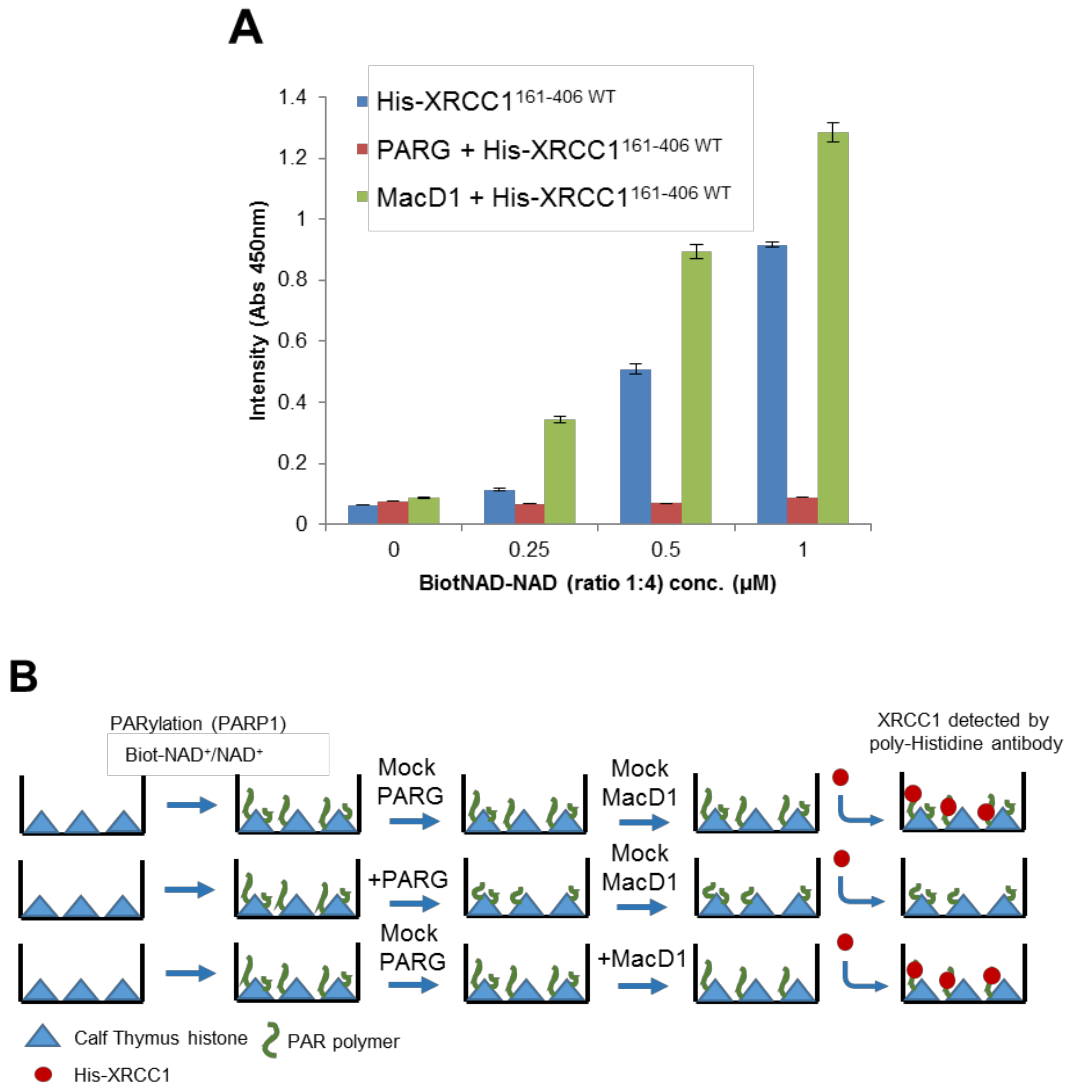

Figure 7.18. The phosphate-binding pocket of the first BRCT domain of XRCC1 binds PAR polymer and not MAR. **A.** Calf thymus histone was ribosylated in the presence of NAD^+ -biotinylated NAD^+ mixture (4 to 1 ratio) with PARP1. The ribosylated proteins were either mock-treated or treated with PARG/MACROD1 then, incubated with $XRCC1^{161\text{-}406\ WT}$. The binding of $XRCC1^{161\text{-}406\ WT}$ was detected as described in Materials and Methods (2.5.2.1). Data are the mean (±1SD) of three independent experiments. (MACROD1 is denoted as MacD1 on the figure) **B.** Schematic representation of the experiment.

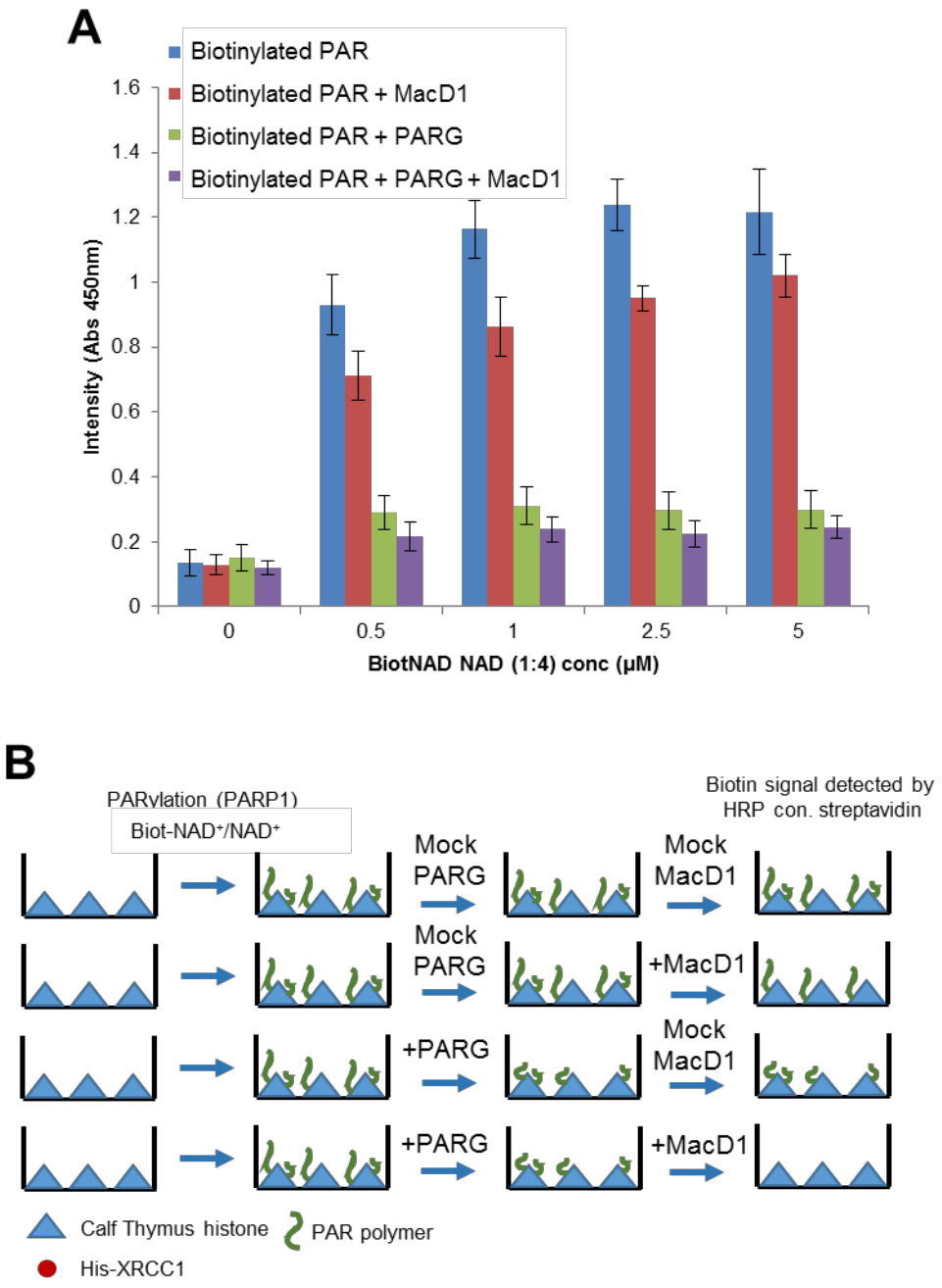

Figure 7.19. The phospho-binding pocket of the first BRCT domain of XRCC1 binds PAR polymer and not MAR. **A.** Calf thymus histone was ribosylated in the presence of NAD^+ -biotinylated NAD^+ mixture (4 to 1 ratio) with PARP1. The ribosylated proteins were either mock-treated or treated with PARG or/and MACROD1 (MacD1 on the figure). The level of ribosylation before and after PARG treatment was detected using HRP-conjugated streptavidin. Data are the mean (±1SD) of three independent experiments. **B.** Schematic representation of the experiment.

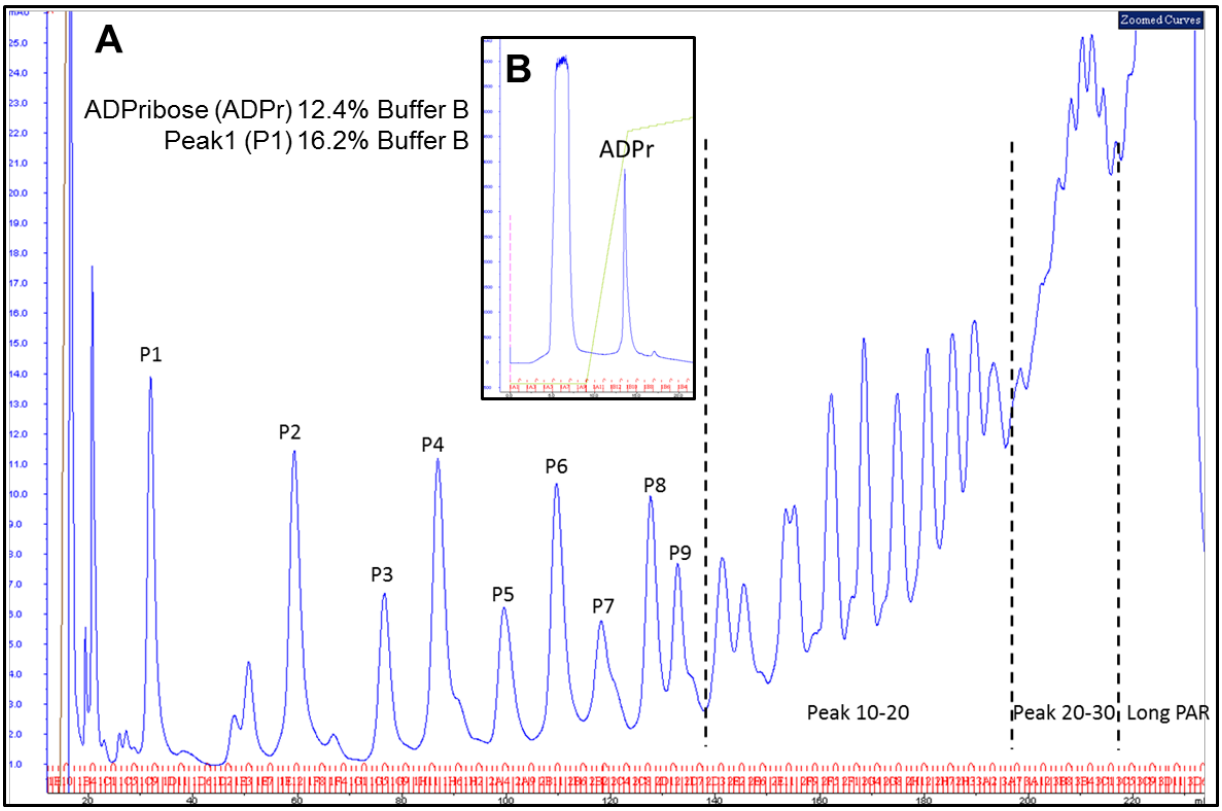

Figure 7.20. Preparation of PAR polymer with distinct unit length. **A.** Fractions of PAR polymers were eluted from MonoQ ION-exchange column by applying increasing salt concentration. Fractions noted from Peak1 (P1) to Peak9 (P9) were collected individually and desalted. Peak 10-20, Peak 20-30 and Long PAR were also collected and desalted as described in Materials and methods. **B.** Mono(ADP-ribose) was also loaded and eluted from MonoQ column and eluted at 12.4% Buffer B.

peak 10-20, and longer PAR lengths were fractionated by denaturing PAGE on sequencing gels and detected by silver staining as described previously (Malanga et al. 1995) above, revealing that each peak fraction represented a high purity polymer with distinct unit length (Figure 7.21) (Materials and Methods 2.5.3.3). It is worth pointing out that during the preparation of PAR, the polymers are detached from the target proteins by KOH treatment, thus resulting in species of different molecular masses for polymers of the same unit length. These differences might arise from decomposition of the proximal ribose moiety (Kiehlbauch et al. 1993). Whilst confirmation requires mass spectrometry, these data suggest that peak 1 is most likely 2 units Peak 2&3 are 3 units, Peak 4&5 are 4 units, etc.

Next, we employed these polymers in competition assays, to further define the unit lengths that best bind XRCC1 (2.5.2.3). The competition assay was carried out using 15nM-15μM (total ADP-ribose units) of each peak fraction (Figure 7.22). Interestingly, between Peak 6 & 7 (both ~5 units) there is a significant increase in ability to bind XRCC1 in these competition assays. Strangely, Peak 8 did not compete at these concentrations, whereas peaks 10-20 and longer polymers were also effective competitors.

7.3 Discussion

In this chapter I demonstrated that the phosphate binding pocket of the XRCC1 BRCT1 domain a PAR binding module. Pull down assays confirmed that the phosphate-binding pocket is required for interaction between ribosylated PARP1 and XRCC1. To further investigate the interaction on molecular level, I progress to experiments employing recombinant proteins. Slot blot experiments in which wild type or mutant $XRCC1^{161\text{-}406}$ was absorbed onto nitrocellulose membrane suggested that binding to auto-ribosylated PARP1 was not greatly impacted by mutation of the phosphate binding pocket. However, I obtained contradictory results if PAR polymer was absorbed onto nylon membrane and was incubated with either wild type or His-$XRCC1^{161\text{-}406\ RK}$. One possible explanation was that there are two modes of XRCC1 binding, one that is dependent on the phosphate-binding pocket and one that is not, with the former evident only at low levels of PARylation.

To test this, I developed and employed a more quantitative ELISA plate-binding assay. This proved successful. The first set of experiments revealed a significant difference in XRCC1 binding between wild type and mutant XRCC1 harbouring a mutated phosphate-

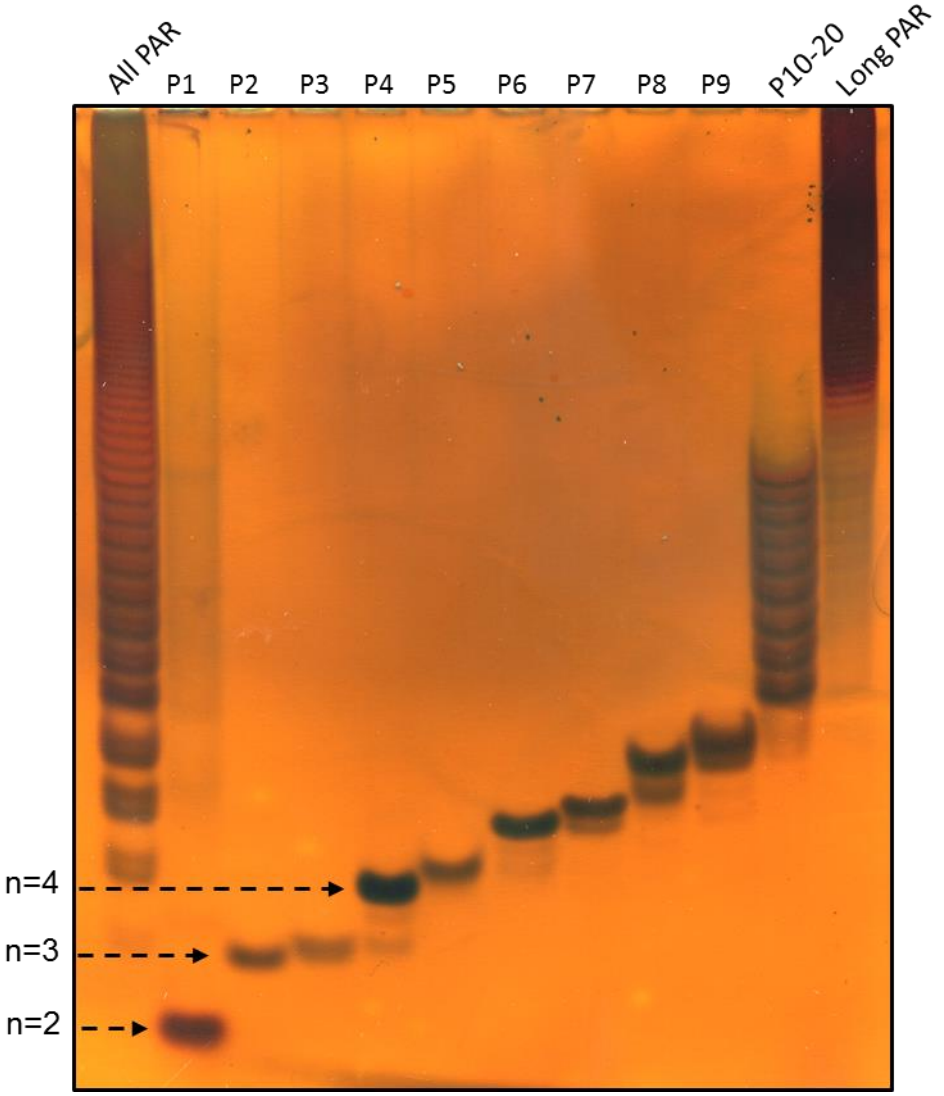

Figure 7.21. Acrylamide gel electrophoresis of purified PAR polymer fractions. PAR fractions were prepared in a final concentration of 0.3μM and diluted in loading buffer (40% urea, 4mM EDTA, 0.02% bromophenol blue, and 0.02% xylene cyanol). The samples were run at a constant power of 15W, then the gel was fixed in 50% ethanol and 5% acetic acid solution for 2 hours. The gel was stained with Pierce® Color Silver Stain Kit (Thermo Scientific).

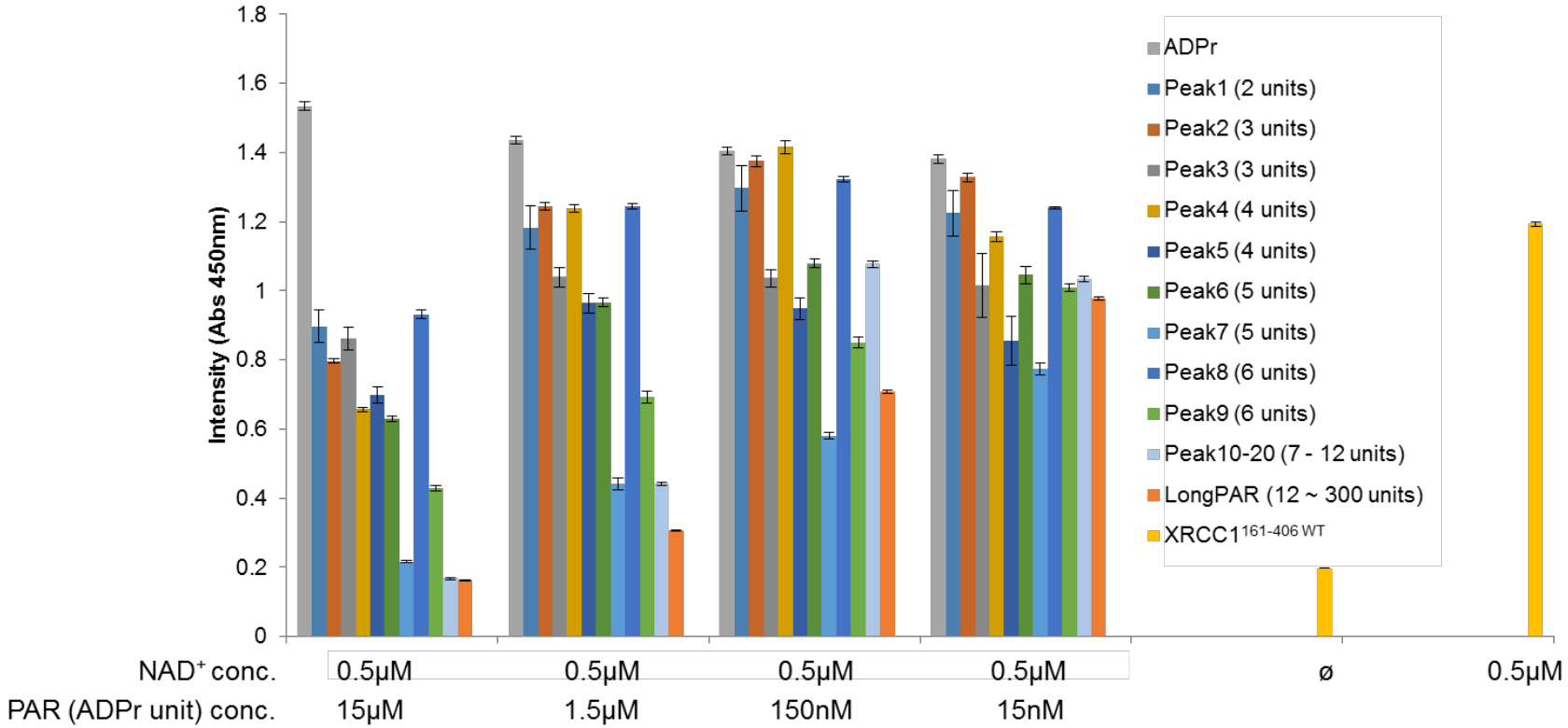

Figure 7.22. PAR competition is dependent on PAR polymer length. Calf thymus histone was ribosylated with 0.5μM NAD+ in the presence of PARP1. XRCC1161-406 fragment was incubated with various concentration (15μM, 1.5μM, 0.15μM and 15nM) of PAR polymers from different fractions, therefore, supposedly with different unit length for 30 min on ice before it was added to the wells. Data are the mean (±1SD) of three independent experiments.

binding pocket, at lower concentrations of NAD^+. However, at higher NAD^+ concentrations, the mutant $XRCC1^{161-406}$ also bound PAR. Formerly, the PAR binding ability was assigned to a degenerate basic patch at the C-terminal part of the first BRCT domain (Pleschke et al. 2000). It is thus possible that this domain bind PAR by itself at high concentrations of polymer, but that this region normally functions cooperatively with the phosphate-binding pocket at lower PAR concentrations. For example, perhaps PAR polymer of an appropriate size can occupy both the phosphate-binding pocket and the degenerate basic patch. However, it is worth noting that in cell based assays loss of the phosphate binding pocket is sufficient to ablate XRCC1 recruitment to damage sites and also XRCC1 functionality, suggesting that at physiological levels of PARylation the phosphate-binding pocket is essential (Breslin et al. 2015).

Consistent with the idea that there may be two cooperative PAR binding sites in the XRCC1 BRCT1 domain, by purifying PAR polymers of distinct unit length it was possible to demonstrate that longer polymers bind to XRCC1 more efficiently, in competition assays. Competition assays also revealed that the phosphate-binding pocket requires at least two ADPr units for binding, since mono(ADP-ribose) (MAR) was not sufficient to compete with ribosylated PARP1 even at >1000x molar excess. This hypothesis was supported by experiments employing PARG or MacD1 to degrade MAR or PAR, in conjunction with biotinylated NAD^+ to measure levels of ADP-ribosylation. These experiments revealed that XRCC1 binding to ribosylated histone protein is sensitive to pre-treatment with PARG, which degrades PAR, but not MacD1, which degrades MAR. Pre-treatment with both enzymes ablated XRCC1 binding.

In future, another way to test the idea that there are two cooperative PAR-binding sites in the BRCT1 domain might be to express and purify a mutant XRCC1 lacking both putative PAR-binding domains. To date, however, this has not proved possible. In addition, X-ray crystallography or NMR may prove useful, especially in conjunction with the PAR molecules of distinct size that I have purified.

CHAPTER EIGHT

Summary and Discussion

8.1 Overall Perspective

During my PhD, I was fortunate enough to be involved in two exciting projects. The first project was focusing on determining X-ray crystal structures of TDP2 in complex with small molecule inhibitors, in order to aid an ongoing structure-based drug design approach. My second project focused on biochemical characterisation of the interaction between the single-strand break repair protein XRCC1 and poly(ADP-ribose).

8.2 Project I: Small molecule inhibitors of TDP2

8.2.1 TDP2 as a potential drug target

TDP2, also known as TTRAP, was identified in a screen for new tyrosyl-DNA phosphodiesterase activities (Cortes Ledesma et al. 2009). Enzyme assays with a substrate harbouring a 5'-tyrosyl moiety, covalently attached to DNA, clearly demonstrated that the protein TDP2 possessed a phosphodiesterase activity capable of removing this type of adduct. Further experiments proved that TDP2 can also liberate trapped covalent complexes of Topoisomerase 2 (TOP2) from damaged DNA termini; thereby indicating that TDP2 is an important factor for DNA double-strand break repair in vertebrate cells (Zeng et al. 2011) and functions as part of the Non-Homologous End-Joining pathway (NHEJ) (Gomez-Herreros et al. 2013). Cells in which TDP2 is either depleted (siRNA) or deleted are highly sensitive to the TOP2-poison etoposide; suggesting that inhibitors of TDP2 may produce elevated levels of cell death in cancer cells, treated with both agents (Cortes Ledesma et al. 2009).

Recent findings have raised the possibility that inhibitors of TDP2 may also have utility as a monotherapy. For example, it was demonstrated that TDP2 protects transcriptional programmes, which require activation by TOP2 (Gomez-Herreros et al. 2013) and that in hormone-driven cancer cell models, depletion of TDP2 inhibited the transcription of certain genes, even at endogenous levels of abortive TOP2 activity. Therefore, hypothetically, a monotherapy using TDP2 inhibitors to treat hormone-driven cancer cells would also lead to elevated levels of cell death. Interestingly, the enzyme activity of TDP2 is also required for host-pathogen interactions in some members of the picornavirus family (human papilloma virus, rhinovirus; (Virgen-Slane et al. 2012). Rhinoviruses are the most common viral infectious agents, and are the cause of the common cold, as well as meningitis, encephalitis and hepatitis. Targeting TDP2 could therefore offer an excellent

target, for combating viral infections and would open new avenues in anti-picornavirus drug development.

8.2.2 High-throughput Screening (HTS): assay development and compound screening

In order to find promising small molecule inhibitors of TDP2, a high-throughput screening (HTS) method was needed. Thomson and co-workers at the Paterson Institute of Cancer Research, University of Manchester developed a chromogenic assay that allowed the screening of a compound library in a HTS fashion (Thomson et al. 2013). The substrate molecule employed was 4-nitrophenyl phenylphosphonate (NPPP). Approximately 100 000 compounds were screened in the HTS screen, with several toxoflavins and a singleton deazaflavin identified as 'hits' (Raoof et al. 2013). Although the toxoflavins showed clear structure activity relationships (SAR) for inhibition of TDP2, their redox liability hindered any further exploration of this chemical scaffold. However, the deazaflavin singleton was expanded into a promising series of compound, which also showed clear SAR through a secondary screen. The deazaflavin series had no observable redox liability, had good EC_{50} values and showed clear dose dependent inhibition of TDP2 enzyme activity.

8.2.3 Structural studies

In order to have a better understanding of how these compounds inhibit TDP2 enzyme activity, and to aid a structure-based drug design approach, structural studies were undertaken. My involvement with the project started by optimising the expression and purification of recombinant TDP2, in order to obtain multi-milligrams of purified protein to facilitate crystallographic screens (Chapter 4). The expression of recombinant proteins can be challenging, and to maximise the yields of soluble protein, it is often more effective to generate expression constructs that encode compact structural units (domains) of a protein, rather than the full length protein. Protein domains are also generally more amenable to crystallisation. Crystallising the catalytic domain of human TDP2 in complex with a small molecule inhibitor would provide the molecular information necessary for a structure-guided drug discovery approach.

Despite the early difficulties of expressing hTDP2, I managed to obtain a sufficient amount of soluble material, which could then be used in crystallisation trials (as described in chapter 5). These experiments produced crystals diffracting to a resolution of 3.2Å.

However, despite purifying and crystallising Selenomethionine-derivatised protein it was not possible to phase the native diffraction data, due to the poor quality of the data collected for the derivative.

At this point of my PhD, two reports describing X-ray crystal structures of the catalytic domains of TDP2 (from different species) were published (Schellenberg et al. 2012, Shi et al. 2012). We reasoned that we could then simply use the structures as search models in molecular replacement, in order to solve the structure of TDP2 in complex with an inhibitory compound. It was therefore not necessary to continue with an experimental-phasing approach (MIR, MAD, SAD etc).

In collaboration with the Cancer Research UK Manchester Institute (formerly the Paterson Institute) at the University of Manchester, our laboratory was provided with several compounds with IC_{50} values in the low micromolar range. These inhibitors were first tested in a ThermoFluor assay, in order to verify binding to TDP2. From this early series of deazaflavin compounds, C14005, a phenylmethanesulfonamide derivative showed strong binding characteristics, producing a large shift in melting temperature midpoint. This compound then was taken forward into a gel-based assay in order to confirm its inhibitory effect, as compound binding does not necessarily correlate with inhibition. However, the solubility of this compound was particularly low, in both organic and aqueous buffers, limiting the amount of compound that could be combined with TDP2, to take forward into crystallisation trials. Whilst putative co-crystals were obtained, they only diffracted to low resolution.

I was provided with a second compound C14297 that had improved solubility: a tetrazole derivative supplied by the Sussex Drug Discovery Centre at the University of Sussex. As before the compound was tested in both binding and activity assays. Once confirmed in both assay formats, it was taken forward into crystallisation trials in complex with the catalytic domain of hTDP2. The resultant crystals provided a usable diffraction dataset and the structure of the protein/compound complex was solved at a resolution of 3.1Å by molecular replacement using the structure of the catalytic domain of murine TDP2 (mTDP2-CAT) as a search model (Schellenberg et al. 2012). Importantly, this structure provided the first information about the binding mode of the deazaflavin class of TDP2 inhibitor, and revealed it to be radically different to that predicted by molecular docking

studies. However, due to the low resolution of diffraction data, and the generally poor behaviour of hTDP2-CAT^{C161S} crystals, we sought an improved experimental system.

It was reasoned that the murine protein might provide such a system. However, through a series of experiments testing compound binding and effects on enzyme activity, it was shown that the deazaflavins did not inhibit the mouse enzyme. Through the experiments and analyses presented in Chapter 4, it became clear that deazaflavin binding is highly dependent on the amino acids found within the divergent region of the M7-loop of the catalytic domain (Schellenberg et al. 2012). It is worth noting here, that any future *in vitro* / *in vivo* experiments using non-human cells, and / or other model organisms will need to take this particular fact into consideration.

Based on comparison of the structure of hTDP2-CAT bound to C14297, with the murine protein, it was possible to design a novel 'surrogate' protein, m2hTDP2-CAT; generated by mutating key amino acid residues of the mouse protein back to their human equivalents, with the view of obtaining co-crystals that diffracted to high resolution.

This approach was highly successful, and I collected two high resolution datasets, one with a tetrazole derivative (C14297), and a second with a hydroxyphenyl-substituted deazaflavin (LEP-0259), validating this experimental approach, and providing the foundation for an ongoing structure-based drug design programme, at the University of Sussex, aiming to develop small molecule inhibitors of TDP2 as therapeutic agents.

By analysis of the structures obtained, we are able to suggest viable alterations to the structure of these compounds, in order to make a more effective, selective inhibitor e.g. substitutions at the R_1 position of the deazaflavin ring system which extend towards the active site, to pick up additional hydrogen bonds; the tartrate molecule trapped at the active site in the LEP-0259 structure provides a useful path in order to directly connect the compound to the active site.

8.2.4 Cell based assays for analysis of TDP2 inhibitors

It is necessary to conduct assays at the cell and organismal level, as part of a drug development programme, in order to obtain important and relevant information about the solubility, absorption, and activity profiles of these compounds in living matter. In chapter 6, I described the initial results from two different phenotypic assay formats: a clonogenic survival assay, and a DNA double-strand break repair assay (γH2AX assay). Both C14297

and LEP-0259 compounds were tested in both assay formats, but neither produced results that indicated an *in vivo* inhibitory effect. As part of future work, these assays could be repeated at higher compound concentrations or with similar compounds with improved solubility profiles. However, any inhibitory effect that is observed in cells, also needs rigorous testing in control experiments as the observed phenotypes could be produced by off-target effects of the inhibitors. Experiments to examine off-target effects could be carried out in cells where TDP2 has been knocked down or deleted. Interestingly, derivatives of deazaflavins have already been described as effective inhibitors of HDM2 – an ubiquitin E3 ligase – that targets p53 for proteolytic degradation (Kitagaki et al. 2008).

8.3 Project II: The phosphate-binding pocket of the BRCT domain of XRCC1 binds poly(ADP-ribose).

It has been long hypothesised that poly(ADP-ribose) synthesis promotes the acceleration of single stranded DNA break repair (SSBR) by recruiting the scaffold protein XRCC1 (El‐Khamisy et al. 2003), however, reports in the literature (Parsons et al. 2005, Abdou et al. 2015) challenged this idea in the last few years.

My project was a part of a collaborative work regarding XRCC1 recruitment by PAR and its effect on SSBR at the biochemical and cellular level. It is known that PAR binding is mediated by the BRCT1 domain of XRCC1 (El‐Khamisy et al. 2003, Okano et al. 2003) and initially it was assigned to a degenerate motif located at the C-terminal part of BRCT1 (Pleschke et al. 2000). However, unpublished results in our lab suggest that this motif is not required for XRCC1 function (unpublished observation). Recently, PAR binding was designated to the phosphate-binding pocket of BRCT1 (Li et al. 2013) formed by residues Arg335 and Lys369. Mutation analysis conducted in the Caldecott lab has shown that mutating both of these residues resulted in a severe delay in repair of SSBs and reduced resistance to genotoxic agents.

In agreement with Li et al, I found that the phospho-binding motif is responsible for PAR binding, However, in contrast to Li et al, my results suggest that mono(ADP-ribose) is not a partner of XRCC1. My experiments employed a truncated XRCC1 construct encoding only the BRCT1 domain of XRCC1 and an N-terminal extension containing the nuclear localization signal (NLS). This truncated XRCC1 construct (denoted as $XRCC1^{161-406}$) was sufficient to confer resistance in EM9 cells to genotoxins (Breslin et al. 2015).

My work on the project started with XRCC1 pull down assays from stably infected EM9 cells with various XRCC1 constructs. These assays demonstrated that the phosphate-binding pocket located on BRCT1 domain is required for the interaction with PARylated PARP1. In order to analyse the interaction between BRCT1 and PAR on a biochemical level, initially, I employed the slot blot assay format, which concluded in contradictory results. A possible explanation maybe that the non-specific PAR binding motif comes into play due to the high level of ribosylation and the long PAR polymer. Subsequently, I started using the plate-binding assay, which turned out to be a very versatile assay format. These assays showed that PAR binding is phosphate-binding pocket dependent, at least at low

substrate NAD^+ concentration. It is in good agreement with experimental results obtained from cells exposed to UVA laser damage or peroxide (H_2O_2), where the recruitment of the mutant XRCC1 is greatly reduced (Breslin et al. 2015) to sites of DNA damage. In biochemical assays, at higher NAD^+ concentration PAR binding of $XRCC1^{161\text{-}406\ RK}$ is still observable, almost to similar extent than that of the wild type protein. A possible reason may be that in these assays at higher NAD^+ concentration the forming polymer is long enough to make contact with the hypothetical C-terminal degenerative motif (Pleschke et al. 2000). This secondary interaction might not be physiologically relevant, as in cells, there is no observable recruitment of the mutant XRCC1.

As PAR is a homopolymer of ADP-ribose units, our initial assumption was that including ADP-ribose in the assay as competitor would ablate BRCT1 binding to ribosylated histones. To our surprise ADP-ribose did not compete in these assays suggesting, that the phosphate-binding pocket requires PAR and not mono(ADP-ribose) (MAR) for efficient binding, which contradicts with the data published by Li et al (Li et al. 2013). Competition assay with PAR and experiments conducted with PAR processing enzymes completely verified this hypothesis. Further experiments with purified PAR with distinct unit length also showed that the interaction between PAR and XRCC1 is dependent on polymer length. It is in great accordance with a report in which the authors were able to quantitatively monitor the binding affinity of XRCC1 for PAR with a kinetic assay using time-resolved fluorescence resonance energy transfer (TR-FRET) (Kim et al. 2015). The authors stated that the BRCT1 domain has a strong preference for PAR chain with length greater than 7 ADPr units. However, in my hands, competition assays suggested that $His\text{-}XRCC1^{161\text{-}406}$ can bind PAR with shorter chain lengths (5 ADPr unit). However, a deeper understanding of this molecular mechanism of XRCC1 binding to PAR will require structural analyses.

The BRCT1 domain of XRCC1 homologues is well conserved in from plants to mammals (Taylor et al. 2002). In *Arabidopsis* the XRCC1 homologue only harbours the central BRCT1 highlighting the importance its central role in DNA repair (Martinez-Macias et al. 2013). We, and others demonstrated that the phospho-binding pocket of BRCT1 is indispensable in recruitment of XRCC1 to damage sites (Breslin et al. 2015, Li et al. 2013). The question arises: why is this particular PAR-binding module the central organizer and operator in SSBr? I.e. the PBZ domain is also capable of recognising ADPr units on the polymer and the biding affinity is dependent on the number of adenine binding sites within single or tandem PBZ motif (Eustermann et al. 2010, Oberoi et al. 2010). Hypothetically,

PBZ domains in tandem would also accelerate the recruitment of XRCC1. A possible explanation maybe that the phospho-binding pocket has a high selectivity towards poly(ADP-ribose), however, it does not bind to mono-ribosylated target proteins. Aprataxin and PNKP like factor (APLF) harbours the PBZ motif, which is responsible for recognising target proteins mainly ribosylated by PARP3 (Rulten et al. 2008, Rulten et al. 2011). PARP3 mainly mono-ribosylates its target proteins, and with APLF, it act together to accelerate nonhomologous end joining. Perhaps, BRCT1 functions as a selector towards single stranded break repair. It would be interesting to see how XRCC1 recruitment to damage sites would be effected in cells if the BRCT1 domain was replaced by single or tandem PBZ motifs. On a biochemical level, my plate binding assays would clearly prove PBZ domain affinity to mono(ADP-ribose), either in competition assays (as in section 7.2.3.4) or binding assays employing PARP3 or PARP1 and PARG (section 7.2.3.5).

8.4 Conclusions

In this chapter, I have provided an overview of the initial steps in the development of a small molecule inhibitor of TDP2 enzymatic activity; which is hoped to find utility in the treatment of cancer, as well as a possible novel anti-viral agent. The structural studies described in this thesis, have provided the molecular details underpinning the binding mode of the deazaflavin series of compounds. Careful examination of the high resolution structures obtained, has already suggested new routes for improving the binding affinity and selectively of future iterations of inhibitors targeted against TDP2. Although, in my hands, the phenotypic cell-based assays indicated no *in vivo* inhibitory activity, we can certainly look positively into the future. In collaboration with the Sussex Drug Discovery Centre (SDDC), a new series of compounds with better solubility characteristics have already been developed, and are currently in co-crystallisation trials. The work described in this thesis, has provided the framework for the identification of a hit-to-lead candidate molecule in the near future.

My project regarding the recruitment of XRCC1 to sites of DNA damage by poly(ADP-ribose) was part of an ongoing work and was published recently (Breslin et al. 2015). Plate binding assays provided good interpretation how PAR binding is mediated on a biochemical level; however, in order to have a deeper understanding of this interaction, we must pursue structural studies. It is reassuring that there are structures of proteins in complex with PAR reported in the literature (Barkauskaite et al. 2013), however, X-ray

crystallography can be difficult and usually requires substantial amount of protein. Other methods such as small angle X-ray scattering (SAXS) and NMR can provide an easier approach on the expense of far less material. As we managed to obtain purified PAR with distinct unit length and preliminary expression trials with BRCT1 of XRCC1 optimised for structural work concluded in success, it is more than possible to obtain structural data of the interaction.

Appendix

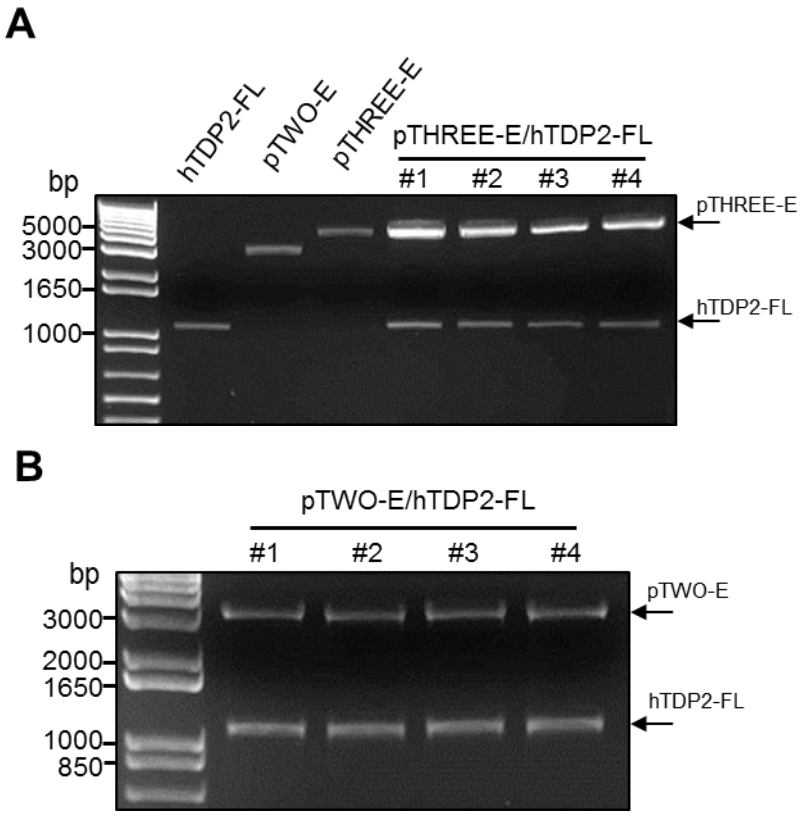

Figure A.1. Cloning of the full-length hTDP2 (hTDP2-FL). The coding DNA was extracted from the original pUC57 vector by NdeI / EcoRI double restriction digest, then cloned into pTHREE-E and pTWO-E expression vectors. **A.** Several pTHREE-E/hTDP2-FL and **B.** pTWO-E/hTDP2-FL clones were subjected to diagnostic digest, then clones #1 and #2 from each construct sent for sequencing. Samples were run in 1% agarose gels, in 1x TAE buffer, and visualised by Ethidium Bromide staining. Ladder: 1Kb Plus DNA ladder. The expected fragment size of hTDP2-FL is 1086bp. Arrows indicate the migrating bands of digested DNA.

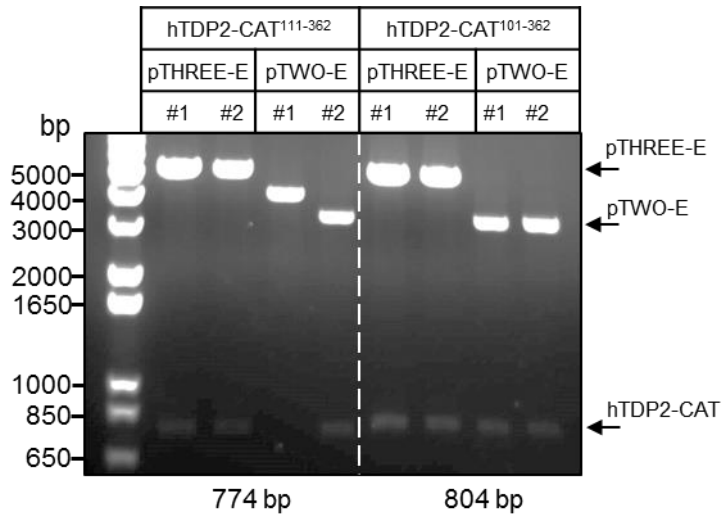

Figure A.2. Cloning of the catalytic domain of hTDP2. Following PCR, the two hTDP2 fragments (hTDP2-CAT$^{111-362}$ and hTDP2-CAT$^{101-362}$) were cloned into pTWO-E and pTHREE-E expression vectors (Table 3.1). Several clones were subjected to diagnostic digest then clones #1 and #2 from each construct sent for sequencing. Samples were run in 1.2% agarose gels, in 1x TAE buffer, and visualised by Ethidium Bromide staining. Ladder: 1Kb Plus DNA ladder. The expected fragment size of hTDP2-CAT$^{111-362}$ is 774bp; hTDP2-CAT$^{101-362}$ is 804bp. Arrows indicate the migrating bands of digested DNA.

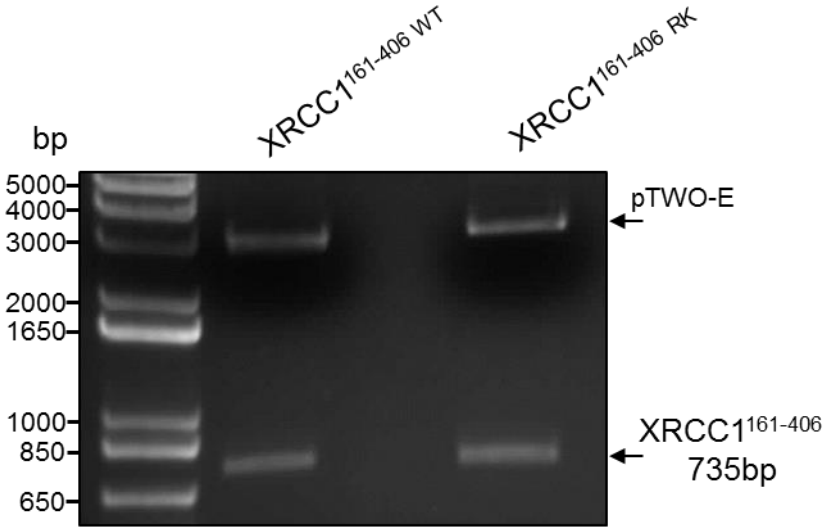

Figure A.3. Cloning of XRCC1$^{161-406}$. Following PCR, the coding DNA was cloned into NdeI / EcoRI double restriction digested pTWO-E. Several clones were subjected to diagnostic digest then clones #1 and #2 from each construct sent for sequencing. Samples were run in 1.2% agarose gels, in 1x TAE buffer, and visualised by Ethidium Bromide staining. Ladder: 1Kb Plus DNA ladder. Expected size of XRCC1$^{161-406}$ is 735bp. Arrows indicate the migrating bands of digested DNA.

References

Abdou, I., G. G. Poirier, M. J. Hendzel and M. Weinfeld (2015). "DNA ligase III acts as a DNA strand break sensor in the cellular orchestration of DNA strand break repair." Nucleic Acids Res **43**(2): 875-892.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

Adhikari, S., S. K. Karmahapatra, T. M. Karve, S. Bandyopadhyay, J. Woodrick, P. V. Manthena, E. Glasgow, S. Byers, T. Saha and A. Uren (2012). "Characterization of magnesium requirement of human 5'-tyrosyl DNA phosphodiesterase mediated reaction." BMC Res Notes **5**: 134.

Aguilera, A. and T. García-Muse (2012). "R Loops: From Transcription Byproducts to Threats to Genome Stability." Molecular Cell **46**(2): 115-124.

Ahnesorg, P., P. Smith and S. P. Jackson (2006). "XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining." Cell **124**(2): 301-313.
Ali, A. A., R. M. Jukes, L. H. Pearl and A. W. Oliver (2009). "Specific recognition of a multiply phosphorylated motif in the DNA repair scaffold XRCC1 by the FHA domain of human PNK." Nucleic Acids Res **37**(5): 1701-1712.

Ali, A. A., G. Timinszky, R. Arribas-Bosacoma, M. Kozlowski, P. O. Hassa, M. Hassler, A. G. Ladurner, L. H. Pearl and A. W. Oliver (2012). "The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks." Nat Struct Mol Biol **19**(7): 685-692.

Ame, J. C., V. Rolli, V. Schreiber, C. Niedergang, F. Apiou, P. Decker, S. Muller, T. Hoger, J. Menissier-de Murcia and G. de Murcia (1999). "PARP-2, A novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase." J Biol Chem **274**(25): 17860-17868.

Ame, J. C., C. Spenlehauer and G. de Murcia (2004). "The PARP superfamily." Bioessays **26**(8): 882-893.

Ames, B. N., M. K. Shigenaga and T. M. Hagen (1993). "Oxidants, antioxidants, and the degenerative diseases of aging." Proc Natl Acad Sci U S A **90**(17): 7915-7922.

Angelis, I. D. and L. Turco (2011). "Caco-2 cells as a model for intestinal absorption." Curr Protoc Toxicol **Chapter 20**: Unit20 26.

Aravind, L., D. D. Leipe and E. V. Koonin (1998). "Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins." Nucleic Acids Res **26**(18): 4205-4213.

Bakkenist, C. J. and M. B. Kastan (2003). "DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation." Nature **421**(6922): 499-506.

Banin, S., L. Moyal, S. Shieh, Y. Taya, C. W. Anderson, L. Chessa, N. I. Smorodinsky, C. Prives, Y. Reiss, Y. Shiloh and Y. Ziv (1998). "Enhanced phosphorylation of p53 by ATM in response to DNA damage." Science **281**(5383): 1674-1677.

Barkauskaite, E., G. Jankevicius, A. G. Ladurner, I. Ahel and G. Timinszky (2013). "The recognition and removal of cellular poly(ADP-ribose) signals." FEBS J **280**(15): 3491-3507.

Bennett, C. B., A. L. Lewis, K. K. Baldwin and M. A. Resnick (1993). "Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid." Proceedings of the National Academy of Sciences **90**(12): 5613-5617.

Berger, J. M., S. J. Gamblin, S. C. Harrison and J. C. Wang (1996). "Structure and mechanism of DNA topoisomerase II." Nature **379**(6562): 225-232.

Boehler, C., L. R. Gauthier, O. Mortusewicz, D. S. Biard, J.-M. Saliou, A. Bresson, S. Sanglier-Cianferani, S. Smith, V. Schreiber, F. Boussin and F. Dantzer (2011). "Poly(ADP-ribose) polymerase 3 (PARP3), a newcomer in cellular response to DNA damage and mitotic progression." Proceedings of the National Academy of Sciences **108**(7): 2783-2788.

Bork, P., K. Hofmann, P. Bucher, A. F. Neuwald, S. F. Altschul and E. V. Koonin (1997). "A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins." Faseb j **11**(1): 68-76.

Boulares, A. H., A. G. Yakovlev, V. Ivanova, B. A. Stoica, G. Wang, S. Iyer and M. Smulson (1999). "Role of poly(ADP-ribose) polymerase (PARP) cleavage in apoptosis. Caspase 3-resistant PARP mutant increases rates of apoptosis in transfected cells." J Biol Chem **274**(33): 22932-22940.

Boulikas, T. (1992). "Poly(ADP-ribose) synthesis and degradation in mammalian nuclei." Anal Biochem **203**(2): 252-258.

Bras, J., I. Alonso, C. Barbot, M. M. Costa, L. Darwent, T. Orme, J. Sequeiros, J. Hardy, P. Coutinho and R. Guerreiro (2015). "Mutations in PNKP cause recessive ataxia with oculomotor apraxia type 4." Am J Hum Genet **96**(3): 474-479.

Braun, S. A., P. L. Panzeter, M. A. Collinge and F. R. Althaus (1994). "Endoglycosidic cleavage of branched polymers by poly(ADP-ribose) glycohydrolase." Eur J Biochem **220**(2): 369-375.

Bredehorst, R., K. Wielckens, P. Adamietz, E. Steinhagen-Thiessen and H. Hilz (1981). "Mono(ADP-ribosyl)ation and poly(ADP-ribosyl)ation of proteins in developing liver and in hepatomas: relation of conjugate subfractions to metabolic competence and proliferation rates." Eur J Biochem **120**(2): 267-274.

Breslin, C., P. Hornyak, A. Ridley, S. L. Rulten, H. Hanzlikova, A. W. Oliver and K. W. Caldecott (2015). "The XRCC1 phosphate-binding pocket binds poly (ADP-ribose) and is required for XRCC1 function." Nucleic Acids Research.

Burzio, L. O., P. T. Riquelme and S. S. Koide (1979). "ADP ribosylation of rat liver nucleosomal core histones." J Biol Chem **254**(8): 3029-3037.

Caldecott, K. W. (2003). "XRCC1 and DNA strand break repair." DNA Repair **2**(9): 955-969.

Caldecott, K. W. (2008). "Single-strand break repair and genetic disease." Nat Rev Genet **9**(8): 619-631.

Caldecott, K. W. (2012). "Tyrosyl DNA phosphodiesterase 2, an enzyme fit for purpose." Nat Struct Mol Biol **19**(12): 1212-1213.

Caldecott, K. W., S. Aoufouchi, P. Johnson and S. Shall (1996). "XRCC1 Polypeptide Interacts with DNA Polymerase β and Possibly Poly (ADP-Ribose) Polymerase, and DNA Ligase III Is a Novel Molecular 'Nick-Sensor' In Vitro." Nucleic Acids Research **24**(22): 4387-4394.

Caldecott, K. W., J. D. Tucker, L. H. Stanker and L. H. Thompson (1995). "Characterization of the XRCC1-DNA ligase III complex in vitro and its absence from mutant hamster cells." Nucleic Acids Res **23**(23): 4836-4843.

Cantoni, O., D. Murray and R. E. Meyn (1987). "Induction and repair of DNA single-strand breaks in EM9 mutant CHO cells treated with hydrogen peroxide." Chem Biol Interact **63**(1): 29-38.

Chambon, P., J. D. Weill and P. Mandel (1963). "Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme." Biochem Biophys Res Commun **11**: 39-43.

Chapman, J. D., J. P. Gagne, G. G. Poirier and D. R. Goodlett (2013). "Mapping PARP-1 auto-ADP-ribosylation sites by liquid chromatography-tandem mass spectrometry." J Proteome Res **12**(4): 1868-1880.

Chayen, N. E. and E. Saridakis (2008). "Protein crystallization: from purified protein to diffraction-quality crystal." Nat Meth **5**(2): 147-153.

Chen, G. L., L. Yang, T. C. Rowe, B. D. Halligan, K. M. Tewey and L. F. Liu (1984). "Nonintercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian DNA topoisomerase II." Journal of Biological Chemistry **259**(21): 13560-13566.

Chen, L., C. J. Nievera, A. Y. Lee and X. Wu (2008). "Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair." J Biol Chem **283**(12): 7713-7720.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Chen, Z., H. Yang and N. P. Pavletich (2008). "Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures." Nature **453**(7194): 489-484.

Chung, T. D., F. H. Drake, K. B. Tan, S. R. Per, S. T. Crooke and C. K. Mirabelli (1989). "Characterization and immunological identification of cDNA clones encoding two human DNA topoisomerase II isozymes." Proc Natl Acad Sci U S A **86**(23): 9431-9435.

Churchill, M. E., J. G. Peak and M. J. Peak (1991). "Correlation between cell survival and DNA single-strand break repair proficiency in the Chinese hamster ovary cell lines AA8 and EM9 irradiated with 365-nm ultraviolet-A radiation." Photochem Photobiol **53**(2): 229-236.

Churchill, M. E., J. G. Peak and M. J. Peak (1991). "Repair of near-visible- and blue-light-induced DNA single-strand breaks by the CHO cell lines AA8 and EM9." Photochem Photobiol **54**(4): 639-644.

Clapperton, J. A., I. A. Manke, D. M. Lowery, T. Ho, L. F. Haire, M. B. Yaffe and S. J. Smerdon (2004). "Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer." Nat Struct Mol Biol **11**(6): 512-518.

Classen, S., S. Olland and J. M. Berger (2003). "Structure of the topoisomerase II ATPase region and its mechanism of inhibition by the chemotherapeutic agent ICRF-187." Proc Natl Acad Sci U S A **100**(19): 10629-10634.

Corbett, K. D. and J. M. Berger (2004). "Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases." Annu Rev Biophys Biomol Struct **33**: 95-118.

Corda, D. and M. Di Girolamo (2003). "Functional aspects of protein mono-ADP-ribosylation." EMBO J **22**(9): 1953-1958.

Cortes Ledesma, F., S. F. El Khamisy, M. C. Zuma, K. Osborn and K. W. Caldecott (2009). "A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage." Nature **461**(7264): 674-678.

Cotner-Gohara, E., I.-K. Kim, A. E. Tomkinson and T. Ellenberger (2008). "Two DNA-binding and Nick Recognition Modules in Human DNA Ligase III." Journal of Biological Chemistry **283**(16): 10764-10772.

Cowell, I. G. and C. A. Austin (2012). "Mechanism of Generation of Therapy Related Leukemia in Response to Anti-Topoisomerase II Agents." International Journal of Environmental Research and Public Health **9**(6): 2075.

D'Amours, D., S. Desnoyers, I. D'Silva and G. G. Poirier (1999). "Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions." Biochem J **342 (Pt 2)**: 249-268.

DaRosa, P. A., Z. Wang, X. Jiang, J. N. Pruneda, F. Cong, R. E. Klevit and W. Xu (2015). "Allosteric activation of the RNF146 ubiquitin ligase by a poly(ADP-ribosyl)ation signal." Nature **517**(7533): 223-226.

Date, H., S. Igarashi, Y. Sano, T. Takahashi, T. Takahashi, H. Takano, S. Tsuji, M. Nishizawa and O. Onodera (2004). "The FHA domain of aprataxin interacts with the C-terminal region of XRCC1." Biochem Biophys Res Commun **325**(4): 1279-1285.

Daugas, E., D. Nochy, L. Ravagnan, M. Loeffler, S. A. Susin, N. Zamzami and G. Kroemer (2000). "Apoptosis-inducing factor (AIF): a ubiquitous mitochondrial oxidoreductase involved in apoptosis." FEBS Letters **476**(3): 118-123.

Davies, D. R., H. Interthal, J. J. Champoux and W. G. J. Hol (2003). "Crystal Structure of a Transition State Mimic for Tdp1 Assembled from Vanadate, DNA, and a Topoisomerase I-Derived Peptide." Chemistry & Biology **10**(2): 139-147.

Demple, B. and M. S. DeMott (2002). "Dynamics and diversions in base excision DNA repair of oxidized abasic lesions." Oncogene **21**(58): 8926-8934.

Derbyshire, D. J. (2002). "Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor." EMBO J. **21**: 3863-3872.

Dong, K. C. and J. M. Berger (2007). "Structural basis for gate-DNA recognition and bending by type IIA topoisomerases." Nature **450**(7173): 1201-1205.

Doogue, M. P. and T. M. Polasek (2013). "The ABCD of clinical pharmacokinetics." Ther Adv Drug Saf **4**(1): 5-7.

Durkacz, B. W., O. Omidiji, D. A. Gray and S. Shall (1980). "(ADP-ribose)n participates in DNA excision repair." Nature **283**(5747): 593-596.

Durocher, D., J. Henckel, A. R. Fersht and S. P. Jackson (1999). "The FHA Domain Is a Modular Phosphopeptide Recognition Motif." Molecular Cell **4**(3): 387-394.

Durocher, D., I. A. Taylor, D. Sarbassova, L. F. Haire, S. L. Westcott, S. P. Jackson, S. J. Smerdon and M. B. Yaffe (2000). "The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms." Mol Cell **6**(5): 1169-1182.

El-Khamisy, S. F. and K. W. Caldecott (2006). "TDP1-dependent DNA single-strand break repair and neurodegeneration." Mutagenesis **21**(4): 219-224.

El-Khamisy, S. F., M. Masutani, H. Suzuki and K. W. Caldecott (2003). "A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage." Nucleic Acids Res **31**(19): 5526-5533.

El-Khamisy, S. F., G. M. Saifi, M. Weinfeld, F. Johansson, T. Helleday, J. R. Lupski and K. W. Caldecott (2005). "Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1." Nature **434**(7029): 108-113.

El‐Khamisy, S. F., M. Masutani, H. Suzuki and K. W. Caldecott (2003). "A requirement for PARP‐1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage." Nucleic Acids Research **31**(19): 5526-5533.

Ellenberger, T. and A. E. Tomkinson (2008). Eukaryotic DNA ligases: Structural and functional insights. Annual Review of Biochemistry. **77:** 313-338.

Elliott, G. and M. Rechsteiner (1975). "Pyridine nucleotide metabolism in mitotic cells." J Cell Physiol **86 Suppl 2**(3 Pt 2): 641-651.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Eustermann, S., C. Brockmann, P. V. Mehrotra, J. C. Yang, D. Loakes, S. C. West, I. Ahel and D. Neuhaus (2010). "Solution structures of the two PBZ domains from human APLF and their interaction with poly(ADP-ribose)." Nat Struct Mol Biol **17**(2): 241-243.

Fan, J., M. Otterlei, H. K. Wong, A. E. Tomkinson and D. M. Wilson (2004). "XRCC1 co‐localizes and physically interacts with PCNA." Nucleic Acids Research **32**(7): 2193-2201.

Fass, D., C. E. Bogden and J. M. Berger (1999). "Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands." Nat Struct Mol Biol **6**(4): 322-326.

Frosina, G., P. Fortini, O. Rossi, F. Carrozzino, G. Raspaglio, L. S. Cox, D. P. Lane, A. Abbondandolo and E. Dogliotti (1996). "Two pathways for base excision repair in mammalian cells." J Biol Chem **271**(16): 9573-9578.

Gao, R., M. J. Schellenberg, S. Y. Huang, M. Abdelmalak, C. Marchand, K. C. Nitiss, J. L. Nitiss, R. S. Williams and Y. Pommier (2014). "Proteolytic degradation of topoisomerase II (Top2) enables the processing of Top2.DNA and Top2.RNA covalent complexes by tyrosyl-DNA-phosphodiesterase 2 (TDP2)." J Biol Chem **289**(26): 17960-17969.

Gary, R., K. Kim, H. L. Cornelius, M. S. Park and Y. Matsumoto (1999). "Proliferating cell nuclear antigen facilitates excision in long-patch base excision repair." J Biol Chem **274**(7): 4354-4363.

Gibson, B. A. and W. L. Kraus (2012). "New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs." Nat Rev Mol Cell Biol **13**(7): 411-424.

Gobeil, S., C. C. Boucher, D. Nadeau and G. G. Poirier (2001). "Characterization of the necrotic cleavage of poly(ADP-ribose) polymerase (PARP-1): implication of lysosomal proteases." Cell Death Differ **8**(6): 588-594.

Golderer, G. and P. Grobner (1991). "ADP-ribosylation of core histones and their acetylated subspecies." Biochem J **277 (Pt 3)**: 607-610.

Gomez-Herreros, F., R. Romero-Granados, Z. Zeng, A. Alvarez-Quilon, C. Quintero, L. Ju, L. Umans, L. Vermeire, D. Huylebroeck, K. W. Caldecott and F. Cortes-Ledesma (2013). "TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo." PLoS Genet **9**(3): e1003226.

Gomez-Herreros, F., J. H. Schuurs-Hoeijmakers, M. McCormack, M. T. Greally, S. Rulten, R. Romero-Granados, T. J. Counihan, E. Chaila, J. Conroy, S. Ennis, N. Delanty, F. Cortes-Ledesma, A. P. de Brouwer, G. L. Cavalleri, S. F. El-Khamisy, B. B. de Vries and K. W. Caldecott (2014). "TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function." Nat Genet **46**(5): 516-521.

Goodwin, P. M., P. J. Lewis, M. I. Davies, C. J. Skidmore and S. Shall (1978). "The effect of gamma radiation and neocarzinostatin on NAD and ATP levels in mouse leukaemia cells." Biochim Biophys Acta **543**(4): 576-582.

Gottlieb, T. M. and S. P. Jackson (1993). "The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen." Cell **72**(1): 131-142.

Grawunder, U., M. Wilm, X. Wu, P. Kulesza, T. E. Wilson, M. Mann and M. R. Lieber (1997). "Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells." Nature **388**(6641): 492-495.

Gryk, M. R., A. Marintchev, M. W. Maciejewski, A. Robertson, S. H. Wilson and G. P. Mullen (2002). "Mapping of the interaction interface of DNA polymerase beta with XRCC1." Structure **10**(12): 1709-1720.

Ha, H. C. and S. H. Snyder (1999). "Poly(ADP-ribose) polymerase is a mediator of necrotic cell death by ATP depletion." Proceedings of the National Academy of Sciences **96**(24): 13978-13982.

Haffner, M. C., A. M. De Marzo, A. K. Meeker, W. G. Nelson and S. Yegnasubramanian (2011). "Transcription-induced DNA double strand breaks: both oncogenic force and potential therapeutic target?" Clin Cancer Res **17**(12): 3858-3864.

Hartlerode, A. J. and R. Scully (2009). "Mechanisms of double-strand break repair in somatic mammalian cells." Biochem J **423**(2): 157-168.

Hartsuiker, E., M. J. Neale and A. M. Carr (2009). "Distinct Requirements for the Rad32Mre11 Nuclease and Ctp1CtIP in the Removal of Covalently Bound Topoisomerase I and II from DNA." Molecular Cell **33**(1): 117-123.

He, F., K. Tsuda, M. Takahashi, K. Kuwasako, T. Terada, M. Shirouzu, S. Watanabe, T. Kigawa, N. Kobayashi, P. Guntert, S. Yokoyama and Y. Muto (2012). "Structural insight into the interaction of ADP-ribose with the PARP WWE domains." FEBS Lett **586**(21): 3858-3864.

Helleday, T. (2011). "The underlying mechanism for the PARP and BRCA synthetic lethality: Clearing up the misunderstandings." Molecular Oncology **5**(4): 387-393.

Heyer, W.-D. (2004). "Recombination: Holliday Junction Resolution and Crossover Formation." Current Biology **14**(2): R56-R58.

Hofmann, K. and P. Bucher (1995). "The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors." Trends Biochem Sci **20**(9): 347-349.

Huang, S., B. Li, M. D. Gray, J. Oshima, I. S. Mian and J. Campisi (1998). "The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease." Nat Genet **20**(2): 114-116.

Huen, M. S., R. Grant, I. Manke, K. Minn, X. Yu, M. B. Yaffe and J. Chen (2007). "RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly." Cell **131**(5): 901-914.

Iles, N., S. Rulten, S. F. El-Khamisy and K. W. Caldecott (2007). "APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks." Mol Cell Biol **27**(10): 3793-3803.

Interthal, H., H. J. Chen, T. E. Kehl-Fie, J. Zotzmann, J. B. Leppard and J. J. Champoux (2005). "SCAN1 mutant Tdp1 accumulates the enzyme--DNA intermediate and causes camptothecin hypersensitivity." EMBO J **24**(12): 2224-2233.

Interthal, H., J. J. Pouliot and J. J. Champoux (2001). "The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily." Proc Natl Acad Sci U S A **98**(21): 12009-12014.

Jackson, S. P. and J. Bartek (2009). "The DNA-damage response in human biology and disease." Nature **461**(7267): 1071-1078.

Jankevicius, G., M. Hassler, B. Golia, V. Rybin, M. Zacharias, G. Timinszky and A. G. Ladurner (2013). "A family of macrodomain proteins reverses cellular mono-ADP-ribosylation." Nat Struct Mol Biol **20**(4): 508-514.

Jarvinen, T. A. and E. T. Liu (2006). "Simultaneous amplification of HER-2 (ERBB2) and topoisomerase IIalpha (TOP2A) genes--molecular basis for combination chemotherapy in cancer." Curr Cancer Drug Targets **6**(7): 579-602.

Jaskulski, D., J. K. deRiel, W. E. Mercer, B. Calabretta and R. Baserga (1988). "Inhibition of cellular proliferation by antisense oligodeoxynucleotides to PCNA cyclin." Science **240**(4858): 1544-1546.

Ju, B. G., V. V. Lunyak, V. Perissi, I. Garcia-Bassets, D. W. Rose, C. K. Glass and M. G. Rosenfeld (2006). "A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription." Science **312**(5781): 1798-1802.

Kabsch, W. (2010). "Xds." Acta Crystallographica Section D: Biological Crystallography **66**(2): 125-132.

Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-666.

Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips and V. C. Shore (1960). "Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution." Nature **185**(4711): 422-427.

Kiehlbauch, C. C., N. Aboul-Ela, E. L. Jacobson, D. P. Ringer and M. K. Jacobson (1993). "High resolution fractionation and characterization of ADP-ribose polymers." Anal Biochem **208**(1): 26-34.

Kim, I. K., R. A. Stegeman, C. A. Brosey and T. Ellenberger (2015). "A quantitative assay reveals ligand specificity of the DNA scaffold repair protein XRCC1 and efficient disassembly of complexes of XRCC1 and the poly(ADP-ribose) polymerase 1 by poly(ADP-ribose) glycohydrolase." J Biol Chem **290**(6): 3775-3783.

Kim, M. Y., S. Mauro, N. Gevry, J. T. Lis and W. L. Kraus (2004). "NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1." Cell **119**(6): 803-814.

Kim, M. Y., T. Zhang and W. L. Kraus (2005). "Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD+ into a nuclear signal." Genes Dev **19**(17): 1951-1967.

Kimura, K., M. Saijo, M. Ui and T. Enomoto (1994). "Growth state- and cell cycle-dependent fluctuation in the expression of two forms of DNA topoisomerase II and possible specific modification of the higher molecular weight form in the M phase." J Biol Chem **269**(2): 1173-1176.

Kitagaki, J., K. K. Agama, Y. Pommier, Y. Yang and A. M. Weissman (2008). "Targeting tumor cells expressing p53 with a water-soluble inhibitor of Hdm2." Mol Cancer Ther **7**(8): 2445-2454.

Koonin, E. V., S. F. Altschul and P. Bork (1996). "Functional motifs." Nat Genet **13**(3): 266-268.

Kreimeyer, A., K. Wielckens, P. Adamietz and H. Hilz (1984). "DNA repair-associated ADP-ribosylation in vivo. Modification of histone H1 differs from that of the principal acceptor proteins." J Biol Chem **259**(2): 890-896.

Krishnakumar, R., M. J. Gamble, K. M. Frizzell, J. G. Berrocal, M. Kininis and W. L. Kraus (2008). "Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes." Science **319**(5864): 819-821.

Krogh, B. O. and S. Shuman (2000). "Catalytic Mechanism of DNA Topoisomerase IB." Molecular Cell **5**(6): 1035-1041.

Krokan, H. E., R. Standal and G. Slupphaug (1997). "DNA glycosylases in the base excision repair of DNA." Biochem J **325 (Pt 1)**: 1-16.

Kubota, Y., R. A. Nash, A. Klungland, P. Schär, D. E. Barnes and T. Lindahl (1996). "Reconstitution of DNA base excision-repair with purified human proteins: Interaction between DNA polymerase β and the XRCC1 protein." EMBO Journal **15**(23): 6662-6670.

Kuzminov, A. (2001). "Single-strand interruptions in replicating chromosomes cause double-strand breaks." Proc Natl Acad Sci U S A **98**(15): 8241-8246.

Langelier, M. F., J. L. Planck, S. Roy and J. M. Pascal (2012). "Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1." Science **336**(6082): 728-732.

Langelier, M. F., A. A. Riccio and J. M. Pascal (2014). "PARP-2 and PARP-3 are selectively activated by 5' phosphorylated DNA breaks through an allosteric regulatory mechanism shared with PARP-1." Nucleic Acids Res **42**(12): 7762-7775.

Langelier, M. F., K. M. Servent, E. E. Rogers and J. M. Pascal (2008). "A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation." J Biol Chem **283**(7): 4105-4114.

Lee, B.-H., K. Yoshimatsu, A. Maeda, K. Ochiai, M. Morimatsu, K. Araki, M. Ogino, S. Morikawa and J. Arikawa (2003). "Association of the nucleocapsid protein of the Seoul and Hantaan hantaviruses with small ubiquitin-like modifier-1-related molecules." Virus Research **98**(1): 83-91.

Lee, Y., S. Katyal, Y. Li, S. F. El-Khamisy, H. R. Russell, K. W. Caldecott and P. J. McKinnon (2009). "The genesis of cerebellar interneurons and the prevention of neural DNA damage require XRCC1." Nat Neurosci **12**(8): 973-980.

Leslie, A. G. (2006). "The integration of macromolecular diffraction data." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 48-57.

Leung, A. K. (2014). "Poly(ADP-ribose): an organizer of cellular architecture." J Cell Biol **205**(5): 613-619.

Leung, C. C. and J. N. Glover (2011). "BRCT domains: easy as one, two, three." Cell Cycle **10**(15): 2461-2470.

Li, A. P. (2001). "Screening for human ADME/Tox drug properties in drug discovery." Drug Discovery Today **6**(7): 357-366.

Li, C., S. Y. Sun, F. R. Khuri and R. Li (2011). "Pleiotropic functions of EAPII/TTRAP/TDP2: cancer development, chemoresistance and beyond." Cell Cycle **10**(19): 3274-3283.

Li, M., L. Y. Lu, C. Y. Yang, S. Wang and X. Yu (2013). "The FHA and BRCT domains recognize ADP-ribosylation during DNA damage response." Genes Dev **27**(16): 1752-1768.

Li, M. and X. C. Yu (2013). "Function of BRCA1 in the DNA Damage Response Is Mediated by ADP-Ribosylation." Cancer Cell **23**(5): 693-704.

Li, N. and J. Chen (2014). "ADP-ribosylation: activation, recognition, and removal." Mol Cells **37**(1): 9-16.

Lieber, M. R. (2010). "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway." Annu Rev Biochem **79**: 181-211.

Lin, W., J. C. Ame, N. Aboul-Ela, E. L. Jacobson and M. K. Jacobson (1997). "Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase." J Biol Chem **272**(18): 11895-11901.

Lindahl, T. and B. Nyberg (1972). "Rate of depurination of native deoxyribonucleic acid." Biochemistry **11**(19): 3610-3618.

Liu, Y., W. A. Beard, D. D. Shock, R. Prasad, E. W. Hou and S. H. Wilson (2005). "DNA polymerase beta and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair." J Biol Chem **280**(5): 3665-3674.

Lo, M. C., A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas and G. Ellestad (2004). "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery." Anal Biochem **332**(1): 153-159.

Loeffler, P. A., M. J. Cuneo, G. A. Mueller, E. F. DeRose, S. A. Gabel and R. E. London (2011). "Structural studies of the PARP-1 BRCT domain." BMC Struct Biol **11**: 37.

Loizou, J. I., S. F. El-Khamisy, A. Zlatanou, D. J. Moore, D. W. Chan, J. Qin, S. Sarno, F. Meggio, L. A. Pinna and K. W. Caldecott (2004). "The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks." Cell **117**(1): 17-28.

London, R. E. (2015). "The structural basis of XRCC1-mediated DNA repair." DNA Repair (Amst) **30**: 90-103.

Long, B. H., S. T. Musial and M. G. Brattain (1985). "Single- and double-strand DNA breakage and repair in human lung adenocarcinoma cells exposed to etoposide and teniposide." Cancer Res **45**(7): 3106-3112.

Lounnas, V., T. Ritschel, J. Kelder, R. McGuire, R. P. Bywater and N. Foloppe (2013). "Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery." Comput Struct Biotechnol J **5**: e201302011.

Luo, X. and W. L. Kraus (2012). "On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1." Genes Dev **26**(5): 417-432.

Lyu, Y. L., C. P. Lin, A. M. Azarova, L. Cai, J. C. Wang and L. F. Liu (2006). "Role of topoisomerase IIbeta in the expression of developmentally regulated genes." Mol Cell Biol **26**(21): 7929-7941.

Ma, Y., U. Pannicke, K. Schwarz and M. R. Lieber (2002). "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination." Cell **108**(6): 781-794.

Mahajan, A., C. Yuan, H. Lee, E. S.-W. Chen, P.-Y. Wu and M.-D. Tsai (2008). "Structure and Function of the Phosphothreonine-Specific FHA Domain." Science Signaling **1**(51): re12-re12.

Mahajan, K. N., S. A. Nick McElhinny, B. S. Mitchell and D. A. Ramsden (2002). "Association of DNA Polymerase μ (pol μ) with Ku and Ligase IV: Role for pol μ in End-Joining Double-Strand Break Repair." Molecular and Cellular Biology **22**(14): 5194-5202.

Malanga, M., S. Bachmann, P. L. Panzeter, B. Zweifel and F. R. Althaus (1995). "Poly(ADP-ribose) quantification at the femtomole level in mammalian cells." Anal Biochem **228**(2): 245-251.

Mani, R. S., F. Karimi-Busheri, M. Fanta, K. W. Caldecott, C. E. Cass and M. Weinfeld (2004). "Biophysical characterization of human XRCC1 and its binding to damaged and undamaged DNA." Biochemistry **43**(51): 16505-16514.

Mao, Y., S. D. Desai, C. Y. Ting, J. Hwang and L. F. Liu (2001). "26 S proteasome-mediated degradation of topoisomerase II cleavable complexes." J Biol Chem **276**(44): 40652-40658.

Marsischky, G. T., B. A. Wilson and R. J. Collier (1995). "Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins." J Biol Chem **270**(7): 3247-3254.

Martinez-Macias, M. I., D. Cordoba-Canero, R. R. Ariza and T. Roldan-Arjona (2013). "The DNA repair protein XRCC1 functions in the plant DNA demethylation pathway by stimulating cytosine methylation (5-meC) excision, gap tailoring, and DNA ligation." J Biol Chem **288**(8): 5496-5505.

Masson, M., C. Niedergang, V. Schreiber, S. Muller, J. Menissier-de Murcia and G. de Murcia (1998). "XRCC1 Is Specifically Associated with Poly(ADP-Ribose) Polymerase and Negatively Regulates Its Activity following DNA Damage." Molecular and Cellular Biology **18**(6): 3563-3571.

Mateyak, M. K. and T. G. Kinzy (2013). "ADP-ribosylation of translation elongation factor 2 by diphtheria toxin in yeast inhibits translation and cell separation." J Biol Chem **288**(34): 24647-24655.

Matsuoka, S., M. Huang and S. J. Elledge (1998). "Linkage of ATM to Cell Cycle Regulation by the Chk2 Protein Kinase." Science **282**(5395): 1893-1897.

Meibohm, B. and H. Derendorf (1997). "Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modelling." Int J Clin Pharmacol Ther **35**(10): 401-413.

Menissier de Murcia, J., M. Ricoul, L. Tartier, C. Niedergang, A. Huber, F. Dantzer, V. Schreiber, J. C. Ame, A. Dierich, M. LeMeur, L. Sabatier, P. Chambon and G. de Murcia (2003). "Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse." EMBO J **22**(9): 2255-2263.

Minaga, T. and E. Kun (1983). "Probable helical conformation of poly(ADP-ribose). The effect of cations on spectral properties." J Biol Chem **258**(9): 5726-5730.

Miwa, M., N. Saikawa, Z. Yamaizumi, S. Nishimura and T. Sugimura (1979). "Structure of poly(adenosine diphosphate ribose): identification of 2'-[1"-ribosyl-2"-(or 3"-)(1"'-ribosyl)]adenosine-5',5",5"'-tris(phosphate) as a branch linkage." Proc Natl Acad Sci U S A **76**(2): 595-599.

Miwa, M. and T. Sugimura (1971). "Splitting of the Ribose-Ribose Linkage of Poly(Adenosine Diphosphate-Ribose) by a Calf Thymus Extract." Journal of Biological Chemistry **246**(20): 6362-6364.

Moreira, M. C., C. Barbot, N. Tachi, N. Kozuka, E. Uchida, T. Gibson, P. Mendonca, M. Costa, J. Barros, T. Yanagisawa, M. Watanabe, Y. Ikeda, M. Aoki, T. Nagata, P. Coutinho, J. Sequeiros and M. Koenig (2001). "The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin." Nat Genet **29**(2): 189-193.

Mueller-Dieckmann, C., S. Kernstock, M. Lisurek, J. P. von Kries, F. Haag, M. S. Weiss and F. Koch-Nolte (2006). "The structure of human ADP-ribosylhydrolase 3 (ARH3) provides insights into the reversibility of protein ADP-ribosylation." Proc Natl Acad Sci U S A **103**(41): 15026-15031.

Murai, J., S. Y. Huang, B. B. Das, A. Renaud, Y. Zhang, J. H. Doroshow, J. Ji, S. Takeda and Y. Pommier (2012). "Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors." Cancer Res **72**(21): 5588-5599.

Murray, J. M., M. Tavassoli, R. al-Harithy, K. S. Sheldrick, A. R. Lehmann, A. M. Carr and F. Z. Watts (1994). "Structural and functional conservation of the human homolog of the Schizosaccharomyces pombe rad2 gene, which is required for chromosome segregation and recovery from DNA damage." Mol Cell Biol **14**(7): 4878-4888.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 355-367.

Niere, M., M. Mashimo, L. Agledal, C. Dolle, A. Kasamatsu, J. Kato, J. Moss and M. Ziegler (2012). "ADP-ribosylhydrolase 3 (ARH3), not poly(ADP-ribose) glycohydrolase (PARG) isoforms, is responsible for degradation of mitochondrial matrix-associated poly(ADP-ribose)." J Biol Chem **287**(20): 16088-16102.

Nitiss, J. L. (2009). "DNA topoisomerase II and its growing repertoire of biological functions." Nat Rev Cancer **9**(5): 327-337.

Nitiss, J. L. (2009). "Targeting DNA topoisomerase II in cancer chemotherapy." Nat Rev Cancer **9**(5): 338-350.

Nnakwe, C. C., M. Altaf, J. Cote and S. J. Kron (2009). "Dissection of Rad9 BRCT domain function in the mitotic checkpoint response to telomere uncapping." DNA Repair (Amst) **8**(12): 1452-1461.

Oberoi, J., M. W. Richards, S. Crumpler, N. Brown, J. Blagg and R. Bayliss (2010). "Structural basis of poly(ADP-ribose) recognition by the multizinc binding domain of checkpoint with forkhead-associated and RING Domains (CHFR)." J Biol Chem **285**(50): 39348-39358.

Ogata, N., K. Ueda and O. Hayaishi (1980). "ADP-ribosylation of histone H2B. Identification of glutamic acid residue 2 as the modification site." J Biol Chem **255**(16): 7610-7615.

Ogata, N., K. Ueda, H. Kagamiyama and O. Hayaishi (1980). "ADP-ribosylation of histone H1. Identification of glutamic acid residues 2, 14, and the COOH-terminal lysine residue as modification sites." J Biol Chem **255**(16): 7616-7620.

Okano, S., L. Lan, K. W. Caldecott, T. Mori and A. Yasui (2003). "Spatial and temporal cellular responses to single-strand breaks in human cells." Mol Cell Biol **23**(11): 3974-3981.

Okazaki, I. and J. Moss (1996). Structure and function of eukaryotic mono-ADP-ribosyltransferases. Reviews of Physiology Biochemistry and Pharmacology, Volume 129, Springer Berlin Heidelberg. **129:** 51-104.

Okazaki, I. J. and J. Moss (1999). "Characterization of glycosylphosphatidylinositiol-anchored, secreted, and intracellular vertebrate mono-ADP-ribosyltransferases." Annu Rev Nutr **19**: 485-509.

Otto, H., P. A. Reche, F. Bazan, K. Dittmar, F. Haag and F. Koch-Nolte (2005). "In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs)." BMC Genomics **6**: 139.

Palmer, I. and P. T. Wingfield (2004). "Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli." Curr Protoc Protein Sci **Chapter 6**: Unit 6 3.

Parsons, J. L., Dianova, II, E. Boswell, M. Weinfeld and G. L. Dianov (2005). "End-damage-specific proteins facilitate recruitment or stability of X-ray cross-complementing protein 1 at the sites of DNA single-strand break repair." FEBS J **272**(22): 5753-5763.

Parsons, J. L., Dianova, II, D. Finch, P. S. Tait, C. E. Strom, T. Helleday and G. L. Dianov (2010). "XRCC1 phosphorylation by CK2 is required for its stability and efficient DNA repair." DNA Repair (Amst) **9**(7): 835-841.

Pei, H., J. S. Yordy, Q. Leng, Q. Zhao, D. K. Watson and R. Li (2003). "EAPII interacts with ETS1 and modulates its transcriptional function." Oncogene **22**(18): 2699-2709.

Pleschke, J. M., H. E. Kleczkowska, M. Strohm and F. R. Althaus (2000). "Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins." J Biol Chem **275**(52): 40974-40980.

Pogozelski, W. K. and T. D. Tullius (1998). "Oxidative Strand Scission of Nucleic Acids: Routes Initiated by Hydrogen Abstraction from the Sugar Moiety." Chemical Reviews **98**(3): 1089-1108.

Pommier, Y., R. E. Schwartz, L. A. Zwelling and K. W. Kohn (1985). "Effects of DNA intercalating agents on topoisomerase II induced DNA strand cleavage in isolated mammalian cell nuclei." Biochemistry **24**(23): 6406-6410.

Pype, S., W. Declercq, A. Ibrahimi, C. Michiels, J. G. Van Rietschoten, N. Dewulf, M. de Boer, P. Vandenabeele, D. Huylebroeck and J. E. Remacle (2000). "TTRAP, a novel protein that associates with CD40, tumor necrosis factor (TNF) receptor-75 and TNF receptor-associated factors (TRAFs), and that inhibits nuclear factor-kappa B activation." J Biol Chem **275**(24): 18586-18593.

Raoof, A., P. Depledge, N. M. Hamilton, N. S. Hamilton, J. R. Hitchin, G. V. Hopkins, A. M. Jordan, L. A. Maguire, A. E. McGonagle, D. P. Mould, M. Rushbrooke, H. F. Small, K. M. Smith, G. J. Thomson, F. Turlais, I. D. Waddell, B. Waszkowycz, A. J. Watson and D. J. Ogilvie (2013). "Toxoflavins and Deazaflavins as the First Reported Selective Small Molecule Inhibitors of Tyrosyl-DNA Phosphodiesterase II." Journal of Medicinal Chemistry **56**(16): 6352-6370.

Richardson, C. and M. Jasin (2000). "Frequent chromosomal translocations induced by DNA double-strand breaks." Nature **405**(6787): 697-700.

Robinson, M. J. and N. Osheroff (1990). "Stabilization of the topoisomerase II-DNA cleavage complex by antineoplastic drugs: inhibition of enzyme-mediated DNA religation by 4'-(9-acridinylamino)methanesulfon-m-anisidide." Biochemistry **29**(10): 2511-2515.

Rodrigues-Lima, F., M. Josephs, M. Katan and B. Cassinat (2001). "Sequence analysis identifies TTRAP, a protein that associates with CD40 and TNF receptor-associated factors, as a member of a superfamily of divalent cation-dependent phosphodiesterases." Biochem Biophys Res Commun **285**(5): 1274-1279.

Rogakou, E. P., D. R. Pilch, A. H. Orr, V. S. Ivanova and W. M. Bonner (1998). "DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139." J Biol Chem **273**(10): 5858-5868.

Rosano, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in Escherichia coli: advances and challenges." Front Microbiol **5**: 172.

Ruf, A., G. de Murcia and G. E. Schulz (1998). "Inhibitor and NAD+ binding to poly(ADP-ribose) polymerase as derived from crystal structures and homology modeling." Biochemistry **37**(11): 3893-3900.

Rulten, S. L. and K. W. Caldecott (2013). "DNA strand break repair and neurodegeneration." DNA Repair (Amst) **12**(8): 558-567.

Rulten, S. L., F. Cortes-Ledesma, L. Guo, N. J. Iles and K. W. Caldecott (2008). "APLF (C2orf13) is a novel component of poly(ADP-ribose) signaling in mammalian cells." Mol Cell Biol **28**(14): 4620-4628.

Rulten, S. L., A. E. Fisher, I. Robert, M. C. Zuma, M. Rouleau, L. Ju, G. Poirier, B. Reina-San-Martin and K. W. Caldecott (2011). "PARP-3 and APLF function together to accelerate nonhomologous end-joining." Mol Cell **41**(1): 33-45.

Rulten, S. L., A. E. O. Fisher, I. Robert, M. C. Zuma, M. Rouleau, L. Ju, G. Poirier, B. Reina-San-Martin and K. W. Caldecott (2011). "PARP-3 and APLF Function Together to Accelerate Nonhomologous End-Joining." Molecular Cell **41**(1): 33-45.

Rupnik, A., N. F. Lowndes and M. Grenon (2010). "MRN and the race to the break." Chromosoma **119**(2): 115-135.

Rybenkov, V. V., C. Ullsperger, A. V. Vologodskii and N. R. Cozzarelli (1997). "Simplification of DNA topology below equilibrium values by type II topoisomerases." Science **277**(5326): 690-693.

Rykhlevskaya, A. I. and S. A. Kuznetsova (2000). "Mono- and Bifunctional DNA Glycosylases Involved in Repairing Oxidatively Damaged DNA." Molecular Biology **34**(6): 860-874.

Sawaya, M. R. (2014). "Methods to refine macromolecular structures in cases of severe diffraction anisotropy." Methods Mol Biol **1091**: 205-214.

Schellenberg, M. J., C. D. Appel, S. Adhikari, P. D. Robertson, D. A. Ramsden and R. S. Williams (2012). "Mechanism of repair of 5'-topoisomerase II-DNA adducts by mammalian tyrosyl-DNA phosphodiesterase 2." Nat Struct Mol Biol **19**(12): 1363-1371.

Schmidt, B. H., N. Osheroff and J. M. Berger (2012). "Structure of a topoisomerase II-DNA-nucleotide complex reveals a new control mechanism for ATPase activity." Nat Struct Mol Biol **19**(11): 1147-1154.

Schreiber, V., F. Dantzer, J.-C. Ame and G. de Murcia (2006). "Poly(ADP-ribose): novel functions for an old molecule." Nat Rev Mol Cell Biol **7**(7): 517-528.

Senisterra, G., I. Chau and M. Vedadi (2012). "Thermal denaturation assays in chemical biology." Assay Drug Dev Technol **10**(2): 128-136.

Sharifi, R., R. Morra, C. D. Appel, M. Tallis, B. Chioza, G. Jankevicius, M. A. Simpson, I. Matic, E. Ozkan, B. Golia, M. J. Schellenberg, R. Weston, J. G. Williams, M. N. Rossi, H. Galehdari, J. Krahn, A. Wan, R. C. Trembath, A. H. Crosby, D. Ahel, R. Hay, A. G. Ladurner, G. Timinszky, R. S. Williams and I. Ahel (2013). "Deficiency of terminal ADP-ribose protein glycohydrolase TARG1/C6orf130 in neurodegenerative disease." EMBO J **32**(9): 1225-1237.

Shen, M. R., M. Z. Zdzienicka, H. Mohrenweiser, L. H. Thompson and M. P. Thelen (1998). "Mutations in hamster single-strand break repair gene XRCC1 causing defective DNA repair." Nucleic Acids Res **26**(4): 1032-1037.

Sheng, Z. Z., Y. Q. Zhao and J. F. Huang (2011). "Functional Evolution of BRCT Domains from Binding DNA to Protein." Evol Bioinform Online **7**: 87-97.

Shi, K., K. Kurahashi, R. Gao, S. E. Tsutakawa, J. A. Tainer, Y. Pommier and H. Aihara (2012). "Structural basis for recognition of 5′ -phosphotyrosine adducts by Tdp2." Nat Struct Mol Biol **19**(12): 1372-1377.

Shivji, K. K., M. K. Kenny and R. D. Wood (1992). "Proliferating cell nuclear antigen is required for DNA excision repair." Cell **69**(2): 367-374.

Simsek, D., A. Furda, Y. Gao, J. Artus, E. Brunet, A. K. Hadjantonakis, B. Van Houten, S. Shuman, P. J. McKinnon and M. Jasin (2011). "Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair." Nature **471**(7337): 245-248.

Sissi, C. and M. Palumbo (2009). "Effects of magnesium and related divalent metal ions in topoisomerase structure and function." Nucleic Acids Res **37**(3): 702-711.

Soldani, C. and A. I. Scovassi (2002). "Poly(ADP-ribose) polymerase-1 cleavage during apoptosis: An update." Apoptosis **7**(4): 321-328.

Sousa, F. G., R. Matuo, D. G. Soares, A. E. Escargueil, J. A. P. Henriques, A. K. Larsen and J. Saffi (2012). "PARPs and the DNA damage response." Carcinogenesis **33**(8): 1433-1440.

Stucki, M., J. A. Clapperton, D. Mohammad, M. B. Yaffe, S. J. Smerdon and S. P. Jackson (2005). "MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks." Cell **123**(7): 1213-1226.

Sukhanova, M. V., S. N. Khodyreva, N. A. Lebedeva, R. Prasad, S. H. Wilson and O. I. Lavrik (2005). "Human base excision repair enzymes apurinic/apyrimidinic endonuclease1 (APE1), DNA polymerase β and poly(ADP-ribose) polymerase 1: interplay between strand-displacement DNA synthesis and proofreading exonuclease activity." Nucleic Acids Research **33**(4): 1222-1229.

Suzuki, H., K. Uchida, H. Shima, T. Sato, T. Okamoto, T. Kimura and M. Miwa (1987). "Molecular cloning of cDNA for human poly(ADP-ribose) polymerase and expression of its gene during HL-60 cell differentiation." Biochem Biophys Res Commun **146**(2): 403-409.

Takashima, H., C. F. Boerkoel, J. John, G. M. Saifi, M. A. M. Salih, D. Armstrong, Y. Mao, F. A. Quiocho, B. B. Roa, M. Nakagawa, D. W. Stockton and J. R. Lupski (2002). "Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy." Nat Genet **32**(2): 267-272.

Tan, E. S., K. A. Krukenberg and T. J. Mitchison (2012). "Large-scale preparation and characterization of poly(ADP-ribose) and defined length polymers." Anal Biochem **428**(2): 126-136.

Taylor, R. M., A. Thistlethwaite and K. W. Caldecott (2002). "Central role for the XRCC1 BRCT I domain in mammalian DNA single-strand break repair." Mol Cell Biol **22**(8): 2556-2563.

Taylor, R. M., B. Wickstead, S. Cronin and K. W. Caldecott (1998). "Role of a BRCT domain in the interaction of DNA ligase III-alpha with the DNA repair protein XRCC1." Curr Biol **8**(15): 877-880.

Tebbs, R. S., M. L. Flannery, J. J. Meneses, A. Hartmann, J. D. Tucker, L. H. Thompson, J. E. Cleaver and R. A. Pedersen (1999). "Requirement for the Xrcc1 DNA base excision repair gene during early mouse development." Dev Biol **208**(2): 513-529.

Thompson, L. H., K. W. Brookman, L. E. Dillehay, A. V. Carrano, J. A. Mazrimas, C. L. Mooney and J. L. Minkler (1982). "A CHO-cell strain having hypersensitivity to mutagens, a defect in DNA strand-break repair, and an extraordinary baseline frequency of sister-chromatid exchange." Mutat Res **95**(2-3): 427-440.

Thomson, G., A. Watson, K. Caldecott, O. Denneny, P. Depledge, N. Hamilton, G. Hopkins, A. Jordan, C. Morrow, A. Raoof, I. Waddell and D. Ogilvie (2013). "Generation of assays and antibodies to facilitate the study of human 5′ -tyrosyl DNA phosphodiesterase." Analytical Biochemistry **436**(2): 145-150.

Timinszky, G., S. Till, P. O. Hassa, M. Hothorn, G. Kustatscher, B. Nijmeijer, J. Colombelli, M. Altmeyer, E. H. Stelzer, K. Scheffzek, M. O. Hottiger and A. G. Ladurner (2009). "A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation." Nat Struct Mol Biol **16**(9): 923-929.

Uchida, K., T. Morita, T. Sato, T. Ogura, R. Yamashita, S. Noguchi, H. Suzuki, H. Nyunoya, M. Miwa and T. Sugimura (1987). "Nucleotide sequence of a full-length cDNA for human fibroblast poly(ADP-ribose) polymerase." Biochem Biophys Res Commun **148**(2): 617-622.

Ushiro, H., Y. Yokoyama and Y. Shizuta (1987). "Purification and characterization of poly(ADP-ribose) synthetase from human placenta." Journal of Biological Chemistry **262**(5): 2352-2357.

Uziel, T., Y. Lerenthal, L. Moyal, Y. Andegeko, L. Mittelman and Y. Shiloh (2003). "Requirement of the MRN complex for ATM activation by DNA damage." The EMBO Journal **22**(20): 5612-5621.

Vagin, A. A., R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long and G. N. Murshudov (2004). "REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2184-2195.

van Maanen, J. M. S., J. Retèl, J. de Vries and H. M. Pinedo (1988). "Mechanism of Action of Antitumor Drug Etoposide: A Review." Journal of the National Cancer Institute **80**(19): 1526-1533.

Virgen-Slane, R., J. M. Rozovics, K. D. Fitzgerald, T. Ngo, W. Chou, G. J. van der Heden van Noort, D. V. Filippov, P. D. Gershon and B. L. Semler (2012). "An RNA virus hijacks an incognito function of a DNA repair enzyme." Proceedings of the National Academy of Sciences **109**(36): 14634-14639.

Wang, Y., V. L. Dawson and T. M. Dawson (2009). "Poly(ADP-ribose) signals to mitochondrial AIF: a key event in parthanatos." Exp Neurol **218**(2): 193-202.

Wang, Z., J. P. Gagne, G. G. Poirier and W. Xu (2014). "Crystallographic and biochemical analysis of the mouse poly(ADP-ribose) glycohydrolase." PLoS One **9**(1): e86010.

Wang, Z., G. A. Michaud, Z. Cheng, Y. Zhang, T. R. Hinds, E. Fan, F. Cong and W. Xu (2012). "Recognition of the iso-ADP-ribose moiety in poly(ADP-ribose) by WWE domains suggests a general mechanism for poly(ADP-ribosyl)ation-dependent ubiquitination." Genes Dev **26**(3): 235-240.

Wilson, D. M., M. Takeshita and B. Demple (1997). "Abasic Site Binding by the Human Apurinic Endonuclease, Ape, and Determination of the DNA Contact Sites." Nucleic Acids Research **25**(5): 933-939.

Wilstermann, A. M. and N. Osheroff (2003). "Stabilization of eukaryotic topoisomerase II-DNA cleavage complexes." Curr Top Med Chem **3**(3): 321-338.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Winter, G. (2010). "xia2: an expert system for macromolecular crystallography data reduction." Journal of Applied Crystallography **43**(1): 186-190.

Wu, H.-Y., S. Shyy, J. C. Wang and L. F. Liu (1988). "Transcription generates positively and negatively supercoiled domains in the template." Cell **53**(3): 433-440.

Wyman, C. and R. Kanaar (2006). "DNA double-strand break repair: all's well that ends well." Annu Rev Genet **40**: 363-383.

Yamamoto, H., K. Schoonjans and J. Auwerx (2007). "Sirtuin functions in health and disease." Mol Endocrinol **21**(8): 1745-1755.

Yamane, K., E. Katayama and T. Tsuruo (2000). "The BRCT regions of tumor suppressor BRCA1 and of XRCC1 show DNA end binding activity with a multimerizing feature." Biochemical and Biophysical Research Communications **279**(2): 678-684.

Yang, S. W., A. B. Burgin, Jr., B. N. Huizenga, C. A. Robertson, K. C. Yao and H. A. Nash (1996). "A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases." Proc Natl Acad Sci U S A **93**(21): 11534-11539.

Yang, X., W. Li, E. D. Prescott, S. J. Burden and J. C. Wang (2000). "DNA topoisomerase IIbeta and neural development." Science **287**(5450): 131-134.

Yazdi, P. T., Y. Wang, S. Zhao, N. Patel, E. Y. Lee and J. Qin (2002). "SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint." Genes Dev **16**(5): 571-582.

Yu, J. and L. Zhang (2008). "PUMA, a potent killer with or without p53." Oncogene **27 Suppl 1**: S71-83.

Yu, X., C. C. Chini, M. He, G. Mer and J. Chen (2003). "The BRCT domain is a phospho-protein binding domain." Science **302**: 639-642.

Zeng, Z., F. Cortes-Ledesma, S. F. El Khamisy and K. W. Caldecott (2011). "TDP2/TTRAP is the major 5'-tyrosyl DNA phosphodiesterase activity in vertebrate cells and is critical for cellular resistance to topoisomerase II-induced DNA damage." J Biol Chem **286**(1): 403-409.

Zeng, Z., A. Sharma, L. Ju, J. Murai, L. Umans, L. Vermeire, Y. Pommier, S. Takeda, D. Huylebroeck, K. W. Caldecott and S. F. El-Khamisy (2012). "TDP2 promotes repair of topoisomerase I-mediated DNA damage in the absence of TDP1." Nucleic Acids Res **40**(17): 8371-8380.

Zhang, A., Y. L. Lyu, C. P. Lin, N. Zhou, A. M. Azarova, L. M. Wood and L. F. Liu (2006). "A protease pathway for the repair of topoisomerase II-DNA covalent complexes." J Biol Chem **281**(47): 35997-36003.

Zhang, X., S. Morera, P. A. Bates, P. C. Whitehead, A. I. Coffer, K. Hainbucher, R. A. Nash, M. J. Sternberg, T. Lindahl and P. S. Freemont (1998). "Structure of an XRCC1 BRCT domain: a new protein-protein interaction module." Embo j **17**(21): 6404-6411.

Zhang, X. and F. W. Studier (1997). "Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme." J Mol Biol **269**(1): 10-27.

Zhang, Y., J. Wang, M. Ding and Y. Yu (2013). "Site-specific characterization of the Asp- and Glu-ADP-ribosylated proteome." Nat Methods **10**(10): 981-984.

Nucleic Acids Research Advance Access published June 29, 2015

doi: 10.1093/nar/gkv623

The XRCC1 phosphate-binding pocket binds poly (ADP-ribose) and is required for XRCC1 function

Claire Breslin†, Peter Hornyak†, Andrew Ridley, Stuart L. Rulten, Hana Hanzlikova, Antony W. Oliver and Keith W. Caldecott*

Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9RQ, UK

Received March 31, 2015; Revised June 01, 2015; Accepted June 04, 2015

ABSTRACT

Poly (ADP-ribose) is synthesized at DNA single-strand breaks and can promote the recruitment of the scaffold protein, XRCC1. However, the mechanism and importance of this process has been challenged. To address this issue, we have characterized the mechanism of poly (ADP-ribose) binding by XRCC1 and examined its importance for XRCC1 function. We show that the phosphate-binding pocket in the central BRCT1 domain of XRCC1 is required for selective binding to poly (ADP-ribose) at low levels of ADP-ribosylation, and promotes interaction with cellular PARP1. We also show that the phosphate-binding pocket is required for EGFP-XRCC1 accumulation at DNA damage induced by UVA laser, H_2O_2, and at sites of sub-nuclear PCNA foci, suggesting that poly (ADP-ribose) promotes XRCC1 recruitment both at single-strand breaks globally across the genome and at sites of DNA replication stress. Finally, we show that the phosphate-binding pocket is required following DNA damage for XRCC1-dependent acceleration of DNA single-strand break repair, DNA base excision repair, and cell survival. These data support the hypothesis that poly (ADP-ribose) synthesis promotes XRCC1 recruitment at DNA damage sites and is important for XRCC1 function.

Single-strand breaks (SSBs) are the commonest lesions arising in cells, resulting both directly from disintegration of deoxyribose and indirectly during the excision repair of DNA base damage [reviewed in (1)]. SSBs usually lack conventional 3′-hydroxyl and 5′-phosphate termini, often possessing modifications such as 3′-phosphate or 5′-hydroxyl termini, or fragments of deoxyribose or topoisomerase. If not repaired rapidly, such termini can block progression of RNA or DNA polymerases, disrupting transcription or replication, respectively. The threat posed by SSBs is indicated by the existence of human genetic diseases associated with neurological dysfunction in which single-strand break repair (SSBR) is attenuated (1).

To date, all known SSBR-defective diseases are associated with defects in end processing, the step of repair during which conventional 3′-hydroxyl and 5′-phosphate termini are restored. One critical component of end processing is XRCC1, a molecular scaffold protein that interacts with and recruits, stabilizes, and/or stimulates end processing enzymes and accelerates SSBR ∼5-fold (2,3). The importance of XRCC1 is illustrated by the hypersensitivity of XRCC1-mutant cells to a broad range of genotoxins and to their elevated frequency of chromosome aberrations, genetic deletions and sister chromatid exchanges (4–6). Moreover, mice with conditional deletion of *Xrcc1* in brain recapitulate many of the pathologies associated with loss of SSBR in humans, including cerebellar defects, ataxia and seizures (7).

A number of observations suggest that XRCC1 recruitment at chromosomal SSBs is promoted by poly (ADP-ribose) (PAR) synthesis. First, XRCC1 interacts directly with both PAR and with the poly (ADP-ribose) polymerases PARP1 and PARP2 (8–10). Second, small molecule-mediated inhibition of PAR synthesis, or depletion/deletion of PARP1, greatly reduces the accumulation of XRCC1 at sites of H_2O_2 or UV laser-induced DNA damage (11–15). Third, mutations that disrupt folding of the PAR-binding BRCT1 domain in XRCC1 reduce or ablate XRCC1 accumulation at DNA damage (9,11,15,16). Finally, depletion of PARG, the enzyme responsible for PAR degradation following SSBR, increases both steady state cellular levels of PAR and the accumulation and/or persistence of XRCC1 in sub-nuclear foci before and after DNA damage (17).

Despite these observations, however, several recent reports have challenged the importance of PAR binding for XRCC1 function, instead ascribing XRCC1 recruitment to DNA binding protein partners such as DNA polymerase β (Polβ), polynucleotide kinase/phosphatase (PNKP), and DNA ligase IIIα (Lig3α) (18–22). One reason this uncertainty remains is a lack of clarity concerning the mechanism of PAR binding by XRCC1. PAR binding was first ascribed

*To whom correspondence should be addressed. Tel: +44 1273 877519; Email: k.w.caldecott@sussex.ac.uk
†These authors contributed equally to the paper as first authors.

to a degenerate motif present at the C-terminus of the central BRCT1 domain in XRCC1, comprised of an alternating series of basic/hydrophobic residues and present in numerous other PAR binding proteins (9). However, a recent study instead assigned PAR binding to the phosphate binding pocket present in the BRCT1 domain (16). Not knowing the site of PAR interaction has prevented the generation of point mutations that specifically reduce or ablate PAR binding, and consequently an analysis for their impact on XRCC1 function. Here, we have confirmed the site of PAR binding in XRCC1, enabling us to mutate this site and address directly, for the first time, its importance for XRCC1 cellular function.

MATERIALS AND METHODS

Cell lines

The osteosarcoma cell line U2OS (obtained from the Genome Damage and Stability Centre cell repository) and derivatives of the Chinese hamster ovary (CHO) cell line EM9 (4) were maintained as monolayers in modified Eagle's medium (MEM) or Dulbecco's modified Eagle's medium (DMEM), respectively, supplemented with 10% (vol/vol) foetal calf serum, 100 U/ml penicillin, 2 mM glutamine and 100 μg/ml streptomycin. Expression constructs were introduced into the *XRCC1*-mutant CHO cell line EM9 by Genejuice transfection (Novagen) and stable cell lines prepared by selection in media containing 1.5 mg/ml G418 (Gibco-Invitrogen) for 10–14 days. The cell line U2OS$^{GFP\text{-}XRCC1}$ was generated by transfection of 1×10^6 U2OS cells with 0.5 μg pEGFP-XRCC1 by nucleofection (Lonza kit V) according to the manufacturer's instructions. Twenty four hours after nucleofection, cells were selected in media containing 1 mg/ml G418 for 3 weeks and single clones selected based on their level of GFP expression. One clone, denoted U2OS$^{GFP\text{-}XRCC1}$, was selected for further use.

XRCC1, PARP1 and PCNA expression constructs

To create pCD2EXH^{R335A,K369A} (denoted pcD2EXH^{RK}), encoding human XRCC1-His ^{R335A,K369A}, the *XRCC1* ORF in pCD2EXH (23) was mutated using a QuikChange Mutagenesis Kit (Agilent Technologies) using the oligonucleotides 5′-CCAGAACCCCTTCGCCTCCGAGCTGCGAG-3′ and 5′-TGCCAACACCCCCGCGTACAGCCAGGTCC-3′. All mutated ORFs were confirmed by sequencing. To create pmRFP-XRCC1$^{161–406}$ and pmRFP-XRCC1$^{161–406\ RK}$, encoding human mRFP-XRCC1$^{161–406}$ and mRFP-XRCC1$^{161–406\ RK,}$ respectively, cDNA encoding XRCC1 residues 161–406 was amplified by PCR using the oligonucleotides 5′-AAGAATTCCATGCACCATCACCATCACCATCC GTCCCAGAAGGTGACAGTG-3′ (forward) and 5′-CCCGAATTCTGCAGTCATGGCCCTGCCATGAG GTA-3′ (reverse) and either pCD2EXH or pCD2E-XH^{RK} as template, as appropriate. Note that these constructs also possess an N-terminal histidine tag (underlined). PCR products were verified by sequencing and cloned into the *Eco*R1 site of pmRFP (24). PARP1-pmCherry and PARP1^{E988K}-pmCherry were kind gifts from Gyula Timinszky (Ludwig-Maximilians University, Munich). pCCC-TagRFP is a chromobody-Tag plasmid encoding PCNA-V_HH fused to RFP (Chromotek), which enables detection of endogenous PCNA.

Western blotting

Proteins were fractionated on 8% or gradient SDS-PAGE gels and transferred onto nitrocellulose membrane (GE Healthcare). Membranes were blocked in 5% non-fat dried milk/TBST for 30 min at room temperature or overnight at 4°C. After washing, blocked membranes were incubated with rabbit anti-GFP polyclonal antibody (#2555, Cell Signalling Technology), rabbit anti-phospho (pS485/pT488) XRCC1 polyclonal antibody (1:5000; A300-231A; Bethyl Laboratories, Inc.), mouse anti-PAR (10H) Mab (1:1000), mouse anti-PARP1 Mab (1:1000, MCA1522G, Bio-Rad), mouse anti-Actin Mab (1:2000, A4700, Sigma), mouse anti-XRCC1 (33-2-5) Mab (23), or mouse anti-polyhistidine (His-tag) Mab (Sigma). Anti-mouse and anti-rabbit HRP-conjugated secondary antibodies (Dako Cytomation) were employed at 1:5000, in 5% non-fat milk for 1 h at room temperature. Detection was by ECL (GE Healthcare) and autoradiography.

Transfection and fluorescence imaging

U2OS cells or U2OS$^{GFP\text{-}XRCC1}$ cells were seeded onto coverslips and transfected 1 day later with appropriate constructs using FuGENE 6 transfection reagent according to the manufacturer (Promega). Twenty four hours later, the cells were mock-treated or treated with 10 mM H_2O_2 for 10 min, incubated at 37°C in drug free media for 15 min, washed with phosphate buffered saline (PBS) and then fixed for 10 min in 4% paraformaldehyde in PBS at room temperature. After fixation the cells were washed 2× with PBS, treated with ice-cold methanol/acetone solution for 10 min, washed 2× with PBS and mounted using VECTASHIELD Mounting Media. Images were captured on a Leica SP8 confocal microscope. For EdU labeling of sites of DNA replication, the Click-iT® EdU Alexa Fluor® 647 Imaging Kit from Molecular Probes was used according to manufacturer's instructions. For laser microirradiation, 2×10^5 EM9 cells were seeded onto glass-bottom dishes (Mattek) and transfected with 1 μg of the indicated pmRFP-XRCC1 construct using Genejuice (Milipore). Twenty four hours later, cells were pre-incubated with 10 μg/ml Hoechst 33258 (for micro-irradiation with a 351 nm laser) or Hoechst 34580 (for a 405 nm laser) for 30 min prior to localised micro-irradiation with a 351 nm or 405 nm UV-laser at a dose of 0.22 J/m^2 as previously described (25). Time-lapse images were recorded at the intervals shown after micro-irradiation. For experiments with PARP inhibitor, cells were pre-incubated with either 100 nM Olaparib (Selleckchem) as indicated or with 500 nM Ku58948 (AstraZeneca) 30 min before micro-irradiation.

Clonogenic survival assays

The indicated cells (500) were plated in duplicate in 10 cm dishes and incubated for 4 h at 37°C. Cells were rinsed with

PBS and either mock treated or treated with H_2O_2 (diluted in PBS at the indicated concentration immediately prior to use) or methyl methanosulfonate (MMS) (diluted in complete medium at the indicated concentration immediately prior to use) for 15 min at room temperature (H_2O_2) or 37°C (MMS). After treatment, cells were washed twice with PBS and incubated for 10–14 days in drug-free medium at 37°C to allow formation of macroscopic colonies. Colonies were fixed in ethanol (95%), stained with 1% methylene blue in 70% ethanol and colonies of >50 cells counted. Percentage survival was calculated for each drug concentration using the equation 100 × [average mean colony number (treated plate)/average mean colony number (untreated plate)].

Alkaline single cell agarose gel electrophoresis (alkaline comet assay)

Sub-confluent cell monolayers were trypsinised, diluted to 2×10^5 cells/ml in ice-cold PBS (for H_2O_2 treatment) or complete media (for MMS treatment) immediately prior to treatment, and either mock-treated or treated with 150 μM H_2O_2 (diluted in ice-cold PBS immediately prior to use) for 20 min on ice or with the indicated concentration of MMS (diluted in complete medium) for 15 min at 37°C. Cells were then rinsed in ice-cold PBS and incubated, where appropriate, in fresh drug-free media for the desired repair period at 37°C. Cells (100 per data point) were then analysed by alkaline comet assay as previously described (26) using Comet Assay IV software (Perceptive Instruments).

Expression and purification of His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$

For expression of recombinant XRCC1 proteins, we employed Rosetta^{TM}2 (DE3)pLysS (Merck Millipore) *Escherichia coli* harbouring the expression plasmids pTWO-E-His-XRCC1$^{161–406}$ or pTWO-E-His-XRCC1$^{161–406\ RK}$. pTWO-E is modified from pET-17b, encoding an N-terminal Rhinovirus 3C-cleavable, His_6 affinity tag. For XRCC1 expression, 100 ml LB ampicillin/chloramphenicol media was inoculated with a single bacterial colony and incubated with shaking (220 rpm) at 37°C for 8 h and then stored at 4°C overnight. The next day, 6× 1 l of LB-ampicillin/chloramphenicol media supplemented with antibiotics as above was inoculated with the starter culture (10 ml/l) and again incubated, with shaking, at 37°C until and OD_{600} of 0.8–1.0 was reached, after which protein expression was induced by the addition of 0.2 ml 1 M IPTG/litre for a period of 3 h at 30°C. Cells were harvested by centrifugation and the resulting pellet stored at −20°C. For purification, cell pellets were thawed on ice, resuspended in 50 mM HEPES pH 7.5, 250 mM NaCl, 10 mM imidazole and 1 mM Tris(2-carboxyethyl)phosphine (TCEP), and lysed by sonication on ice for 10 min (10 s on/10 s off) using a large parallel probe at 25% amplitude (Sonics Vibra-Cell, VWR). The lysate was clarified by centrifugation for 50 min at 40 000 × *g* at 4°C and the resulting supernatant added to a 5 ml bed volume of Talon resin (Clontech) in a gravity flow column. After 30 min incubation with the resin, with mixing at 4°C, unbound material was removed by sequential washes (3 × 10 ml) with resuspension buffer. Bound protein was eluted by addition of (2 × 5 ml) elution buffer (50 mM HEPES pH7.5, 250 mM NaCl, 300 mM imidazole and 1 mM TCEP). The eluate was loaded onto a pre-equilibrated (20 mM HEPES pH 7.5, 150 mM NaCl and 1 mM TCEP) 5 ml FF Heparin column (GE Healthcare) and bound material eluted with a linear salt gradient (20 mM HEPES pH 7.5, 1 M NaCl and 1 mM TCEP). Fractions containing XRCC1 were identified by SDS-PAGE, then pooled and concentrated, using Vivaspin 20 (10 000 MWCO) centrifugal concentrators (Sartorius Stedim), to a final concentration of 0.3 mg/ml, and then stored at −80°C until required.

Thermal denaturation and circular dichroism

For thermal denaturation, samples containing 2.0 μM protein and 5× SYPRO Orange (diluted from a 5000 × stock supplied in DMSO; catalogue number S5692, Sigma–Aldrich) were prepared in sample buffer [50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP and 5× Sypro Orange (from Sigma–Aldrich®)]. Denaturation curves were monitored in 96-well PCR plates in a Roche LightCycler 480 II, using 465 and 580 nm filters for excitation and emission wavelengths, respectively. Temperature midpoints (T_m) for each folded to unfolded transition were determined by non-linear regression fitting of a modified Boltzmann model (27) to normalized data in Prism5 (GraphPad Software).

$$Y = (a_n X + b_n) + \frac{(a_d X + b_d) - (a_n X + b_n)}{1 + e^{\frac{T_m - X}{m}}}$$

where: a_n and a_d are the slopes, b_n and b_d the *y*-intercepts, of the native and denatured baselines, respectively. T_m is the melting temperature and *m* a slope factor.

For circular dichroism, spectra were measured at 20°C between the wavelengths 198 and 280 nm in a JASCO J-715 spectropolarimeter attached to a JASCO PTC-384W temperature control system. CD spectra were measured using a 0.1 mm path length cell (Starna Scientific), with protein at a concentration of 54 μM, that had been buffer-exchanged into 10 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP. Spectra were measured using a 0.1 mm path length cell (Starna Scientific) and represent the average of 10 consecutive scans, where the signal from buffer alone has been subtracted.

Poly (ADP-ribose) binding assays

The wells of flat bottomed 96 well PS-microplates (Greiner) were incubated with either 50 μl recombinant histone H1, PARP1 or BSA at 0.1 mg/ml in PBS overnight at 4°C and the wells rinsed (4×) with 0.2 ml 0.1% Triton X100 in PBS. The adsorbed proteins were mock ribosylated in the absence of NAD^+ or ribosylated in the presence of the indicated concentration of NAD^+ (Sigma) in PARP1 reaction buffer (50 mM Tris–HCl pH8, 0.8 mM $MgCl_2$, 1% glycerol and 1.5 mM DTT) containing 40 nM single-stranded oligodeoxyribonucleotide (Eurogentec: 5′-CATATGCCGGAGATCCGCCTCC-3′) and 5 nM PARP1 (recombinant, human, full length) in a final volume of 50 μl at room temp for 30 min. After rinsing (4×) with

50 μl of 0.1% Tween 20 in PBS, 50 μl of His-XRCC1$^{161–406}$ or His-XRCC1$^{161–406\,RK}$ (diluted to 25 nM in 20 mM Tris pH7.5, 130 nM NaCl) was added to the adsorbed proteins and incubated on ice for 30 min. Where indicated, His-XRCC1 proteins were pre-incubated with mono (ADP-ribose) or poly (ADP-ribose)(Trevigen) competitor at the concentrations indicated for 30 min at 4°C, before their addition to the adsorbed proteins. The wells were then rinsed (4×) as above and incubated with 50 μl mouse anti-polyhistidine (His-tag) Mab (Sigma, diluted 1: 3000 in 20 mM Tris pH7.5, 130 nM NaCl) followed by 50 μl HRP-conjugated rabbit anti-mouse IgG (Dako, 1: 5000 in dilution buffer) for 30 min each on ice. After a final wash with 0.1% Tween 20 in PBS, 50 μl of TACS Sapphire (Trevigen) was added to the wells, incubated in the dark for 15 min, stopped by adding 0.2 M HCl, and the absorbance was read at 450 nm.

GFP pull down experiments

U2OS$^{GFP\text{-}XRCC1\text{-}His}$ cells (see above), or U2OS cells 48 h after nucleofection (Amaxa; Lonza, Slough, UK) with 4 μg each of pEGFP-XRCC1$^{161–406}$ or pEGFP-XRCC1$^{161–406\,RK}$ and either pmCherry-PARP1 or pmCherry-PARP1^{E988K} were snap frozen until needed. Cells were then thawed on ice and lysed on ice for 20 min in 0.4 ml/5 × 10^6 cells in 25 mM HEPES (pH 7.8), 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, and including Protease Inhibitor Cocktail and Phosphatase Inhibitor Cocktail 3 (Sigma-Aldrich®, Dorset, UK). Where indicated, the PARP1 inhibitor KU58948 (500 nM) was added to the cell culture medium 1 h prior to cell harvest and/or was included in the cell lysis buffer. Lysed cells were sonicated in a Bioruptor and clarified by centrifugation at 4°C. Unless stated otherwise, all subsequent steps were performed on ice. Forty microliters of the clarified extract was retained on ice as 'input' and 360 μl was mixed with 15 μl (bed volume) of GFP-Trap_A beads (ChromoTek GmbH, Germany) prewashed in 0.5 ml wash buffer (lysis buffer containing 1 mM DDT and 25 mM imidazole). After 1 h on a carousel at 4°C, the GFP-Trap®_A beads were gently pelleted by centrifugation at 2000 × *g* for 2 min. Sixty microliters of the supernatant was retained as 'unbound' material and the pellet was washed three times in wash buffer, with 50 μl of the final wash retained as 'final wash'. Proteins were eluted from the beads by re-suspension in 50 μl 2× Laemmli buffer (250 mM Tris (pH 8.0), 10% SDS, 500 mM DTT, 50% glycerol)), heating for 5 min at 95°C, and centrifugation at 2700 × *g* for 2 min to recover the supernatant.

RESULTS

To further examine the importance of PAR binding for XRCC1 function we first addressed the location of the PAR-binding site. The most evolutionary conserved and functionally important region of XRCC1 is the central BRCT1 domain that mediates binding to PAR (see Figure 1A) (28). PAR-binding by the BRCT1 domain was initially ascribed to a degenerate motif of hydrophobic/basic amino acids that is present in many PAR binding proteins (Figure 1B, dotted red box) (9). However, a different putative PAR-binding motif in BRCT1 was recently reported, comprised of the phosphate-binding pocket common to several other BRCT domains (Figure 1B, solid red boxes and Figure 1C) (16). Within this pocket, Ser328, Arg335 and Lys369 are all predicted to contribute to phosphate binding, based on the structure of other phosphate-binding BRCT domains of this type. Consequently, for subsequent analysis *in vitro*, we expressed and purified both a wild-type histidine-tagged fragment of human XRCC1 spanning the conserved BRCT1 domain (denoted His-XRCC1$^{161–406}$) and a mutant derivative in which both Arg335 and Lys369 were mutated to Ala (denoted His-XRCC1$^{161–406\,RK}$) (Figure 1D, left). We employed both mutations because mutation of R335 alone failed to have any measurable impact on XRCC1 function (data not shown).

Next, to confirm PAR binding by the BRCT1 phosphate-binding pocket we adsorbed PARP1, histone H1, or BSA to microwell plates, mock-ribosylated or ribosylated these proteins with PARP1 in the absence or presence of NAD^+, respectively, and compared their binding to His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\,RK}$, *in vitro* (Figure 1D, right). Wild-type His-XRCC1$^{161–406}$ bound both to adsorbed PARP1 and histone H1, if these proteins were first ribosylated in the presence of 1–50 μM NAD^+, and was fully bound even at the lowest concentration of NAD^+ employed (1 μM) (Figure 2A, blue bars). In contrast, relatively little binding was observed to BSA, irrespective of whether or not it was first incubated with PARP1 and NAD^+, consistent with this protein being a poor substrate for PARP1. More importantly, His-XRCC1$^{161–406\,RK}$ bound ribosylated PARP1 and histone H1 to a much lesser extent, and not at all at the lowest concentration (1 μM) of NAD^+ employed (Figure 2A, red bars). This did not reflect a non-specific impact of the mutations on folding of the BRCT1 domain, because His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\,RK}$ exhibited similar thermal stabilities and circular dichroism spectra (Figure 2B). Importantly, His-XRCC1$^{161–406}$ bound specifically to PAR in these experiments, because it was suppressed by a 8-fold molar excess of ADP-ribose competitor if present as polymer (PAR), but was not suppressed even at 500-fold molar excess if present as ADP-ribose monomer (MAR) (Figure 2C). These data confirm that the phosphate-binding pocket of the XRCC1 BRCT1 domain promotes binding to PAR *in vitro*, particularly at low levels of poly (ADP-ribosylation).

Next, we examined whether PAR binding by the phosphate-binding pocket is physiologically relevant, by comparing wild type and mutant XRCC1 for interaction with cellular PARP1. As expected, full length EGFP-XRCC1 co-precipitated endogenous PARP1 from stably transfected U2OS cells (U2OS$^{GFP\text{-}XRCC1}$ cells; see 'Materials and Methods' section) in a manner that was inhibited by PARP inhibitor (Figure 3A). Similarly, truncated EGFP-XRCC1$^{161–406}$ spanning the BRCT1 domain co-precipitated mCherry-PARP1 in transient co-transfection experiments, but co-precipitated mutant mCherry-PARP1^{E988K} lacking polymerase activity (29,30) to a much lesser extent (Figure 3B). More importantly, EGFP-XRCC1$^{161–406\,RK}$ was also less able to pull-down wild type mCherry-PARP1, confirming that the phosphate-binding pocket promotes interac-

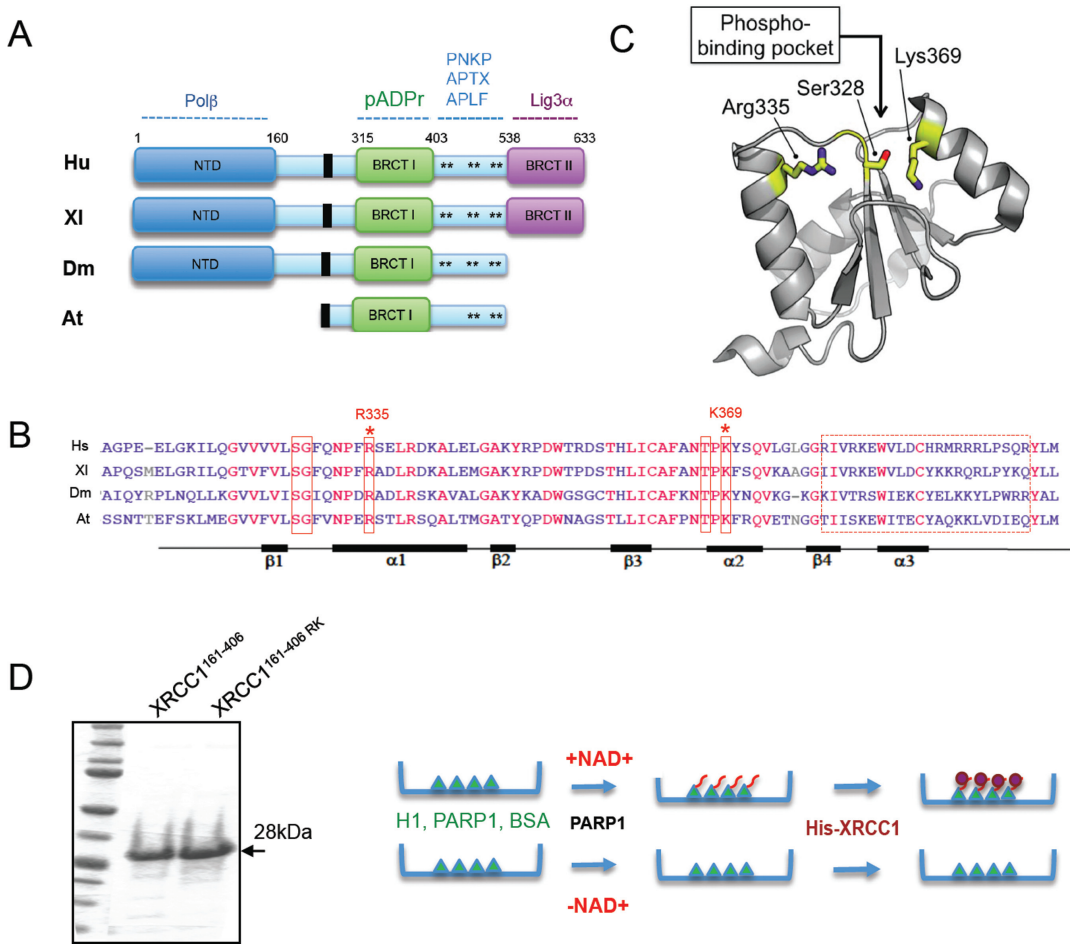

Figure 1. Conservation of the XRCC1 BRCT1 domain and its phosphate-binding pocket. (**A**) Schematic depicting conservation of XRCC1 domains in human (Hs), frog (Xl), fly (Dm), and plant (At) XRCC1. Binding sites for the indicated protein partners are shown. Asterisks denote CK2 phosphorylation sites that mediate FHA-dependent interactions with PNKP, APTX and APLF. Black boxes denoted the nuclear localization signal. (**B**) Alignment of the BRCT1 domain from human, frog, fly and plant. Solid red boxes denote the residues predicted to form the phosphate-binding pocket and the dotted red box denotes the degenerate putative PAR-binding motif identified by Pleschke *et al.* (9). Conserved identical residues are in red. Asterisks denote residues mutated in this study. (**C**) Model of the BRCT1 domain based on the NMR structure (PDN accession code: 2D8M), highlighting the residues predicted to form phosphate-binding pocket. **(D)** Left, purified recombinant His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\,RK}$ proteins, fractionated by SDS-PAGE and stained with coomassie brilliant blue. Right, cartoon of the *in vitro* PAR-binding assay. Proteins were adsorbed to microwell dishes and mock-ribosylated ('-NAD$^+$') or ribosylated ('+NAD$^+$') by PARP1 in absence or presence of NAD$^+$ as indicated. Bound proteins were then incubated with recombinant wild type His-XRCC1$^{161–406}$ or mutant His-XRCC1$^{161–406\,RK}$, and bound XRCC1 detected with anti-His tag antibodies colourmetrically (A_{450}) using HRP-conjugated secondary antibody.

tion with PARP1 (Figure 3B). Consistent with these data, mRFP-XRCC1$^{161–406}$ rapidly accumulated at sites of UVA laser damage at a rate similar to full-length mRFP-XRCC1 and in a manner that was greatly inhibited by PARP inhibitor (500 nM Ku58948), suggesting that the region spanning the BRCT1 domain is sufficient for XRCC1 accumulation at sites of cellular PAR synthesis (Figure 3C and D). Note that we confirmed previously that this concentration of Ku58948 greatly reduces or ablates PAR synthesis in UVA laser tracks (31). In contrast, neither full-length mRFP-XRCC1^{RK} nor mRFP-XRCC1$^{161–406\,RK}$ accumulated at sites of UVA laser damage (Figure 3C and D). Similarly, full-length EGFP-XRCC1^{RK} failed to accumulate in sub-nuclear foci at sites of H_2O_2-induced oxidative stress, confirming that the phosphate-binding pocket is also required for accumulation of EGFP-XRCC1 at this more physiologically relevant source of SSBs (Figure 4A).

XRCC1 has also been reported to colocalise with PCNA in replication foci in human cells, consistent with its proposed role during SSBR at sites of stalled or collapsed replication forks (1,32–35). However, whether XRCC1 accumulation at such sites is also regulated by PAR synthesis is not known. Indeed, the accumulation of EGFP-XRCC1 in sub-nuclear foci with endogenous PCNA, detected by expression of anti-PCNA antibody, was greatly reduced by PARP inhibitor in both early and late S phase cells (Figure 4B, left panels). We confirmed in these experiments that the sites of PCNA and EGFP-XRCC1 colocalisation were sites of DNA replication, by pulse labeling with EdU (Figure 4B, right panels). Importantly, EGFP-XRCC1 accumulation at sites of PCNA accumulation was greatly reduced or ablated by mutation of the phosphate-binding pocket, suggesting that PAR binding is also critical for

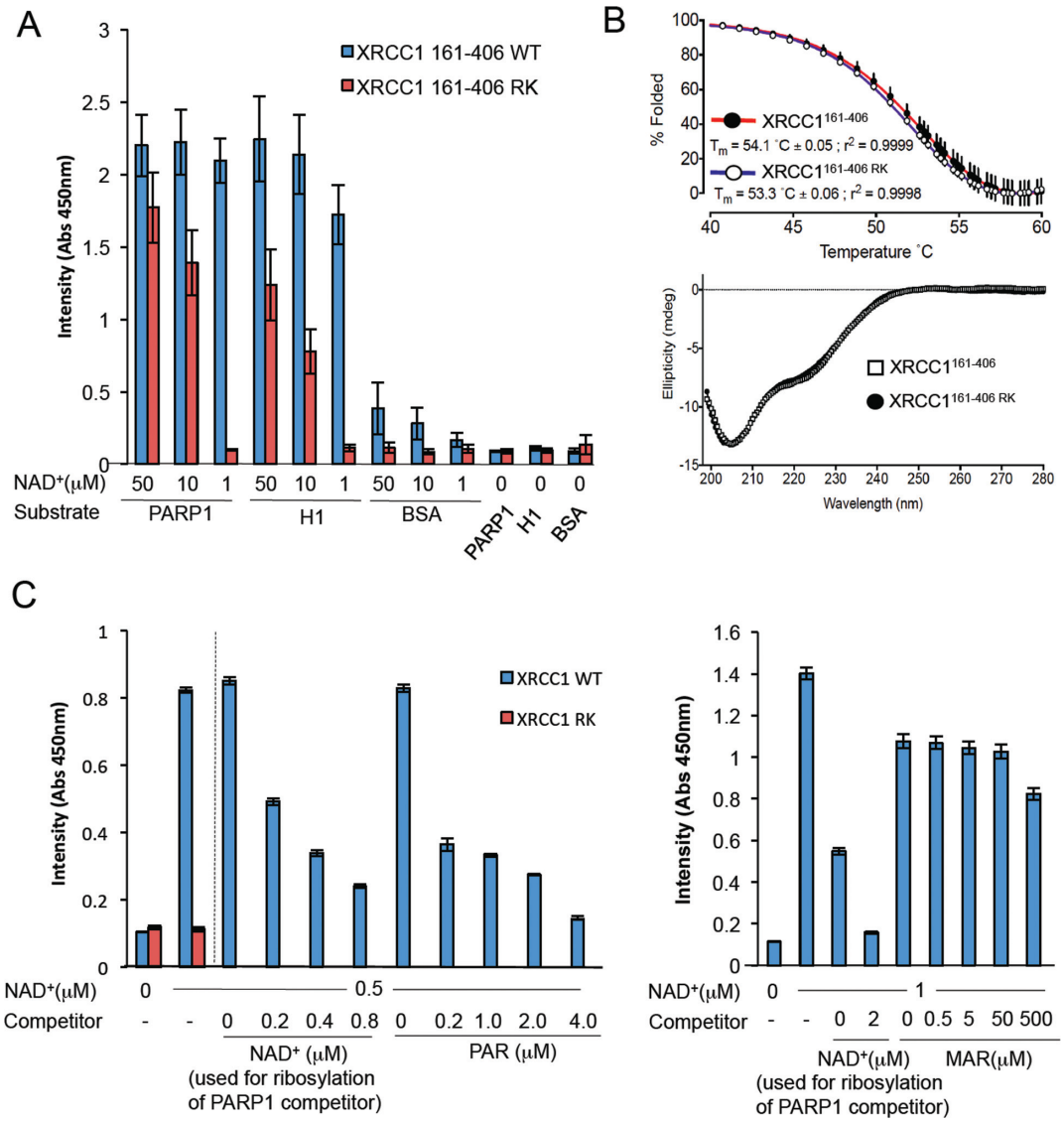

Figure 2. The XRCC1 BRCT1 phosphate-binding pocket binds PAR, *in vitro*. (**A**) Binding of His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$ to the indicated mock-ribosylated (-NAD$^+$) or ribosylated (1–50 μM NAD$^+$) proteins was measured as indicated in Figure 1D. Data are the mean (±1 SD) of at least three experiments. (**B**) Top, thermal stability of recombinant His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$. 2 μM XRCC1 protein was assayed in the presence of SYPRO Orange and unfolding temperatures determined as described in materials and methods. Data are the mean (±1SD) of four independent measurements. Bottom, circular dichroism of His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$. Data are the average of 10 sequential scans, with the spectrum from sample buffer alone subtracted. (**C**) Binding of His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$ to calf thymus histone mock-ribosylated in the absence of NAD+ ('0') or ribosylated in the presence of either 0.5 μM NAD$^+$ (left panel) or 1 μM NAD$^+$ (right panel). Where indicated, XRCC1 binding was measured in the presence of 43 nM PARP1 competitor that was first autoribosylated in the presence of 0, 0.2, 0.4, 0.8 or 2 μM NAD$^+$, as indicated. Alternatively, His-XRCC1$^{161–406}$ and His-XRCC1$^{161–406\ RK}$ binding was measured the presence of the indicated concentration of either poly (ADP-ribose) ('PAR', left) or mono (ADP-ribose) ('MAR', right) competitor. PAR/MAR competitor concentrations are total ADP-ribose units (μM) present as PAR (2–300 subunit lengths) or MAR. Data are the mean (±1SD) of at least three experiments.

the recruitment/accumulation of EGFP-XRCC1 at sites of DNA replication (Figure 4C).

Finally, to address the importance of the phosphate-binding pocket for XRCC1 function, we employed derivatives of *XRCC1*-mutant EM9 cells stably transfected with either empty vector (EM9-V) or with expression vector encoding either full-length human XRCC1-His (EM9-XH) or XRCC1-His^{RK} (EM9-XH^{RK}) (Figure 5A). In contrast to XRCC1-His, XRCC1-His^{RK} was unable to promote cell survival in *XRCC1*-mutant EM9 cells much more than empty vector, following H_2O_2 or MMS treatment (Figure 5B). This was also true in experiments in which we measured rates of chromosomal SSBR using alkaline comet assays, in which XRCC1-His^{RK} again failed to correct the slow rate of DNA strand break repair observed in EM9 cells (Figure 5C).

Collectively, these data demonstrate that the XRCC1 phosphate-binding pocket binds PAR *in vitro* and in cells, promotes XRCC1 accumulation at sites of DNA damage, and is required for XRCC1 cellular function.

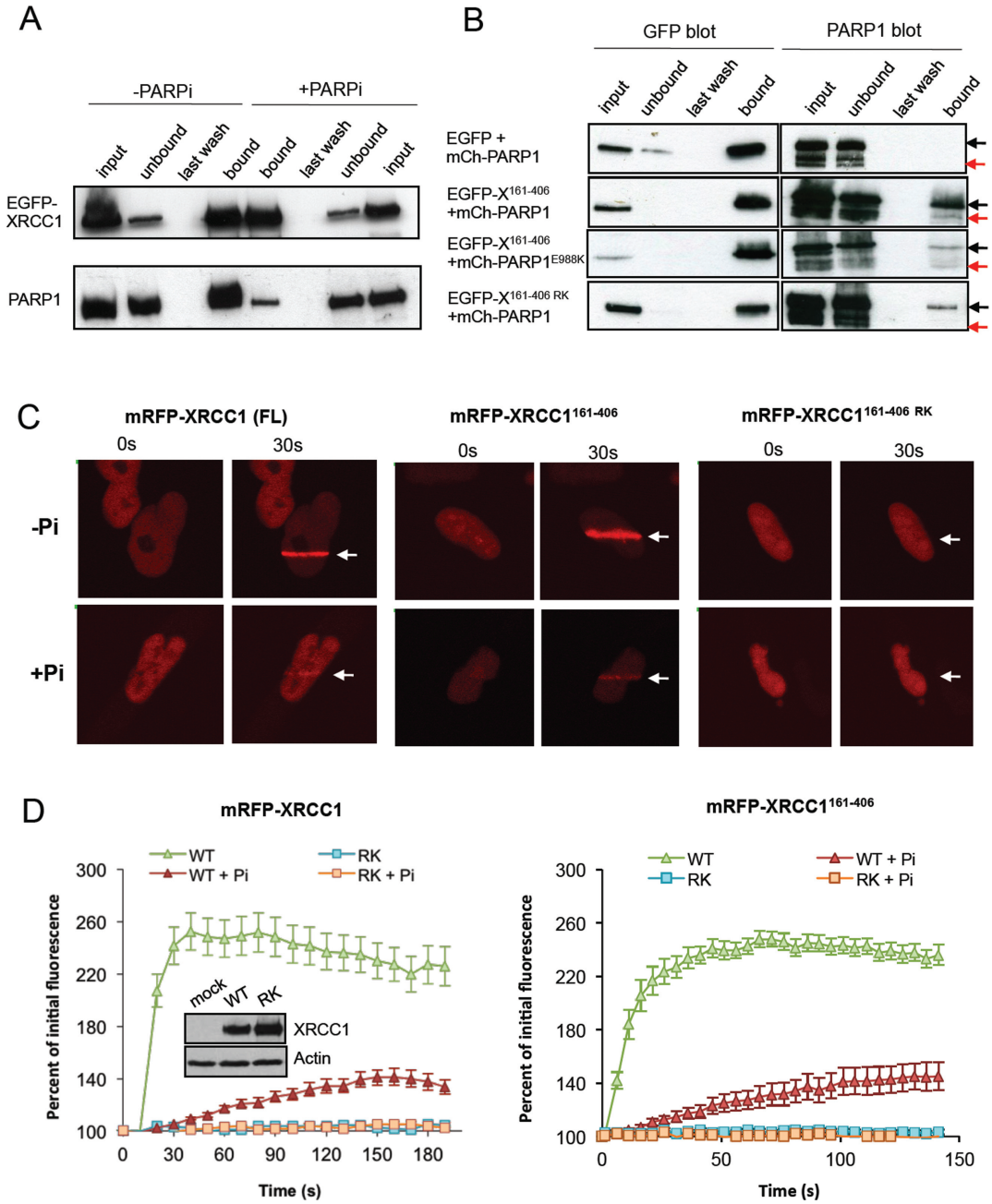

Figure 3. The XRCC1 BRCT1 phosphate-binding pocket mediates PAR-dependent interaction with PARP1 and recruitment at sites of UVA laser induced damage. (**A**) EGFP-XRCC1 was affinity purified from cell extract from U2OS$^{GFP\text{-}XRCC1}$ cells using GFP-Trap beads. Aliquots of the column input, unbound, last wash, and eluate samples were fractionated by SDS-PAGE and immunoblotted with pS485/pT488 anti-XRCC1 polyclonal antibody or anti-PARP1 antibody. Where indicated ('+PARPi'), PARP inhibitor (500 nM Ku-58948) was included in the cell lysis buffer and was present in the cell culture medium for 1 h at 37°C prior to lysis. (**B**) U2OS cells were transiently co-transfected with expression vector encoding either EGFP, EGFP-XRCC1$^{161–406}$, or EGFP-XRCC1$^{161–406\,RK}$ and with expression vector encoding either mCherry-PARP1 or mCherry-PARP1^{E988K}. EGFP-XRCC1 was recovered from whole cell extract and aliquots of column input, unbound, last wash and eluate (bound material) fractionated by SDS-PAGE and immunoblotted with anti-GFP or anti-PARP1 antibody. The position of mCherry-PARP1 and endogenous PARP1 are indicated by black and red arrows, respectively. (**C**) *XRCC1*-mutant EM9 cells were transiently transfected with pmRFP-XRCC1, pmRFP-XRCC1^{RK}, pmRFP-XRCC1$^{161–406}$, or pmRFP-XRCC1$^{161–406\,RK}$ and treated with UVA laser. mRFP fluorescence was measured at the indicated times (seco) following microirradiation in the presence or absence of 500nM PARP inhibitor (Ku-58948). Representative images are shown. (**D**) Left, quantitation of the mRFP-XRCC1 fluorescence proteins at sites of 405 nm UVA laser-induced DNA damage in the above experiments. Inset, pmRFP-XRCC1 and pmRFP-XRCC1^{RK} expression levels in the transfected cells, as measured by immunoblotting with pS485/pT488 anti-XRCC1 polyclonal antibody. Right quantitation of the mRFP-XRCC1^{161_406} fluorescence at sites of 351 nm UVA laser damage. Data is expressed as change in mean fluorescence in ten or more cells per construct ± SEM.

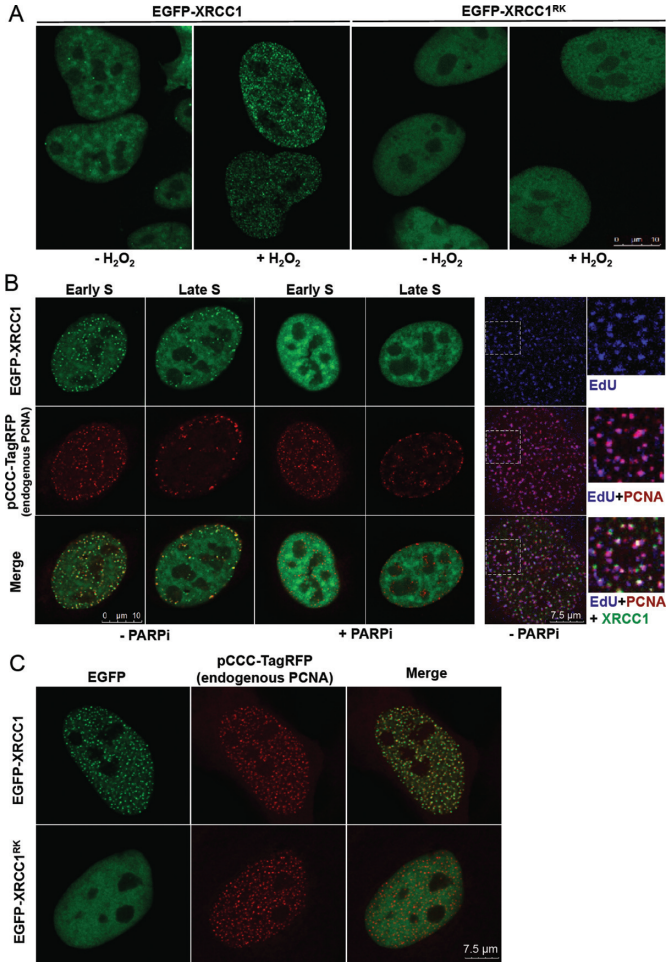

Figure 4. The XRCC1 BRCT1 phosphate-binding pocket is important for XRCC1 accumulation at sites of H_2O_2-induced damage and for colocalization with PCNA. (**A**) U2OS cells were transfected with pEGFP-XRCC1 or pEGFP-XRCC1^{RK}, mock-treated or treated with 10 mM H_2O_2 for 10 min, and after 15 min recovery in drug-free medium fixed and analysed by fluorescence microscopy. (**B**) Left, U2OS$^{GFP\text{-}XRCC1}$ cells were transfected with pCCC-TagRFP to detect endogenous PCNA in the presence or absence of the PARP inhibitor olaparib (100 nM) and analysed as above 24 h later. Right, cells were transfected as above and additionally pulse labelled with EdU (blue) to identify sites of DNA replication. Dotted square denotes the area expanded on the right. (**C**) U2OS cells were co-transfected with either pEGFP-XRCC1^{WT} or pEGFP-XRCC1^{RK} and pCCC-TagRFP plasmid to detect endogenous PCNA. Representative images are shown.

DISCUSSION

The synthesis of poly (ADP-ribose) (PAR) by PARP1 can accelerate SSBR, but the molecular mechanism by which PAR achieves this is unclear (17). One likely role is promoting recruitment of the SSBR scaffold protein, XRCC1 (11–15), although this idea has proved controversial (18–22). To further address this possibility we have clarified the mechanism of PAR binding by XRCC1 and addressed its importance for SSBR and cell survival. PAR binding was initially ascribed to a degenerate motif present at the C-terminus of the central BRCT1 domain in XRCC1, comprised of an alternating series of basic/hydrophobic residues and present in numerous other PAR binding proteins (9). Interestingly, this motif in XRCC1 harbours a common polymorphism at amino acid 399 (arginine/glutamine), which in some epidemiological studies has been implicated in altered predisposition to cancer. However, in cellular assays this polymorphism does not impact measurably on XRCC1 function, suggesting that it does not influence PAR binding (36). Moreover, replacement of five of the basic residues characteristic of this degenerate motif with alanine also fails to impact on XRCC1 function, suggesting that the degenerate motif is not, by itself at least, important for PAR binding (unpublished observations).

Recently, PAR binding by XRCC1 was assigned to a different region of the BRCT1 domain; the highly conserved phosphate binding pocket in (16). In agreement with Li *et al.*, we found that the phosphate-binding pocket interacts directly with PAR. However, in contrast to Li *et al.*, we did not detect binding to mono(ADP-ribose) (MAR) by this motif. Indeed, our competition assays indicate that binding by this motif is highly selective for PAR. We found that the phosphate-binding pocket confers on XRCC1 the ability to bind PAR at low concentrations of polymer, as indicated by its greater impact on PAR binding by XRCC1 at low concentrations of NAD^+, *in vitro*. This might be an advantage at low levels of SSBs such as those arising endogenously in cells, in which PAR polymer might be present at a low concentration and distributed at only a small number of sites across the genome. However, XRCC1 harbouring a mutated phosphate-binding pocket still bound PAR at high concentrations of polymer, albeit to a lesser extent than wild type XRCC1. This may reflect incomplete ablation of PAR binding by the R335A/K369A mutation or, alternatively, weak PAR binding conferred by the degenerate PAR binding motif described above. Nevertheless, mutation of the phosphate-binding pocket greatly reduced mRFP-XRCC1 recruitment at sites of UVA laser-induced damage, and also EGFP-XRCC1 at sites of DNA damage induced by H_2O_2, suggesting that this pocket is critical for accumulation of EGFP-XRCC1 at cellular sites of DNA strand breakage. Interestingly, the impact of mutating the phosphate-binding pocket on XRCC1 accumulation at sites of UVA laser induced damage was greater than incubation with PARP inhibitor. This might reflect incomplete inhibition of PAR synthesis by inhibitor or, alternatively, a low level of protein ribosylation generated prior to incubation with PARP inhibitor.

Mutation of the phosphate-binding pocket also greatly reduced XRCC1 accumulation at sites of PCNA accumulation, suggesting that PAR synthesis also promotes XRCC1 accumulation at sites of damaged replication forks. The latter is consistent with our model for replication-coupled SSBR, in which XRCC1 promotes repair of SSBs either ahead of an approaching fork or after replication fork collapse (35,37). It is also consistent with a role for PARP1 in regulating fork progression in the presence of DNA strand breaks (38–41). However, it is important to note that we have so far only observed XRCC1 accumulation at sites of ongoing DNA replication in cells co-expressing RFP-PCNA or anti-PCNA antibody (data not shown). Consequently, we suggest that both approaches perturb normal PCNA function to some extent, thereby generating

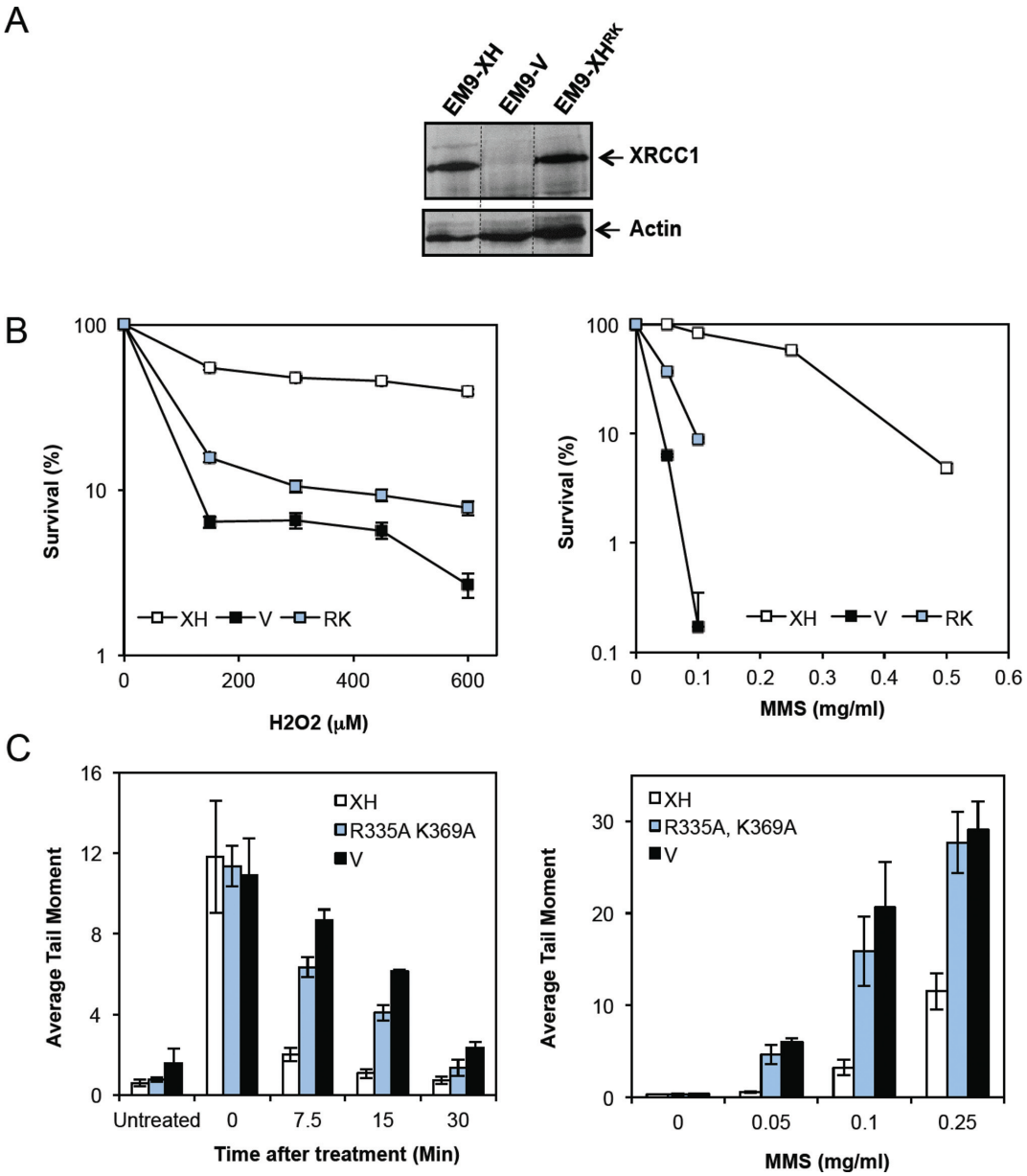

Figure 5. XRCC1-mediated acceleration of SSBR and cell survival requires the XRCC1 BRCT1 phosphate-binding pocket. (**A**) XRCC1 protein expression in *XRCC1-mutant* EM9 cells stably transfected with empty expression vector (EM9-V) or expression vector encoding either XRCC1-His (EM9-XH) or XRCC1-His^{RK} (EM9-XH^{RK}). Cell extracts were fractionated by SDS-PAGE and immunoblotted with anti-XRCC1 Mab (33–2–5) and anti-Actin antibodies. (**B**) Clonogenic survival of *XRCC1*-mutant EM9 cells stably transfected with empty expression vector (EM9-V) or expression vector encoding either XRCC1-His (EM9-XH) or XRCC1-His^{RK} (EM9-XH^{RK}). Cells were treated with the indicated concentrations of H_2O_2 (left) or MMS (right) for 15 min and then in drug free medium for 10–14 days to allow colony formation. Data are the mean (±SEM) of three independent experiments. Where not visible, error bars are smaller than the symbols. (**C**) Chromosomal SSBR rates were measured in the above EM9 cell lines in alkaline comet assays following treatment with 150 μM H_2O_2 for 20 min on ice, followed by recovery in drug-free medium for the indicated time at 37°C, or with the indicated concentration of MMS for 15 min at 37°C to measure the accumulation of SSB intermediates during BER. Data are the mean (±SEM) of three independent experiments.

SSBs and/or other sources of replication stress that trigger PARP1 activation.

Finally, XRCC1 harbouring a mutated phosphate-binding pocket was unable to restore rapid rates of chromosomal SSBR to *XRCC1*-mutant EM9 cells, following treatment with either H_2O_2 or MMS, and only slightly increased cellular resistance to these genotoxins. This work thus highlights the importance of the PAR-binding motif for XRCC1 functionality, both at oxidative breaks induced by H_2O_2 and following MMS-induced DNA alkylation. The latter is particularly intriguing, because MMS-induced SSBs arise as intermediates of DNA base excision repair (BER), suggesting that PAR is important for XRCC1 function during BER. Whereas several reports have suggested that PARP1 is required during BER following DNA alkylation (42,43), others have reported that it is dispensable (44–

46). To reconcile this discrepancy, we previously suggested that PARP1 may be required to detect SSBs arising during BER only if the SSB intermediate becomes uncoupled from the canonical pathway, and/or during replication-coupled SSBR (35,37). However, in the current work, the PAR-binding pocket appeared to be required to accelerate most if not all XRCC1-dependent BER events, as measured by alkaline comet assays following MMS treatment. The extent to which PAR synthesis promotes BER events thus warrants further investigation.

In summary, we confirm that PAR binding is mediated by the phosphate-binding pocket of the XRCC1 BRCT1 domain, and show that the PAR-binding pocket promotes XRCC1 accumulation at DNA damage globally across the genome and at sites of DNA replication stress. We also show that the phosphate-binding pocket is required for acceleration of SSBR by XRCC1, and for XRCC1-dependent cell survival, supporting the hypothesis that poly (ADP-ribose) synthesis is important for XRCC1 recruitment and function.

FUNDING

MRC Programme [G0830 to KWC] and CR-UK Programme [C302/A14532 to A Oliver/L.Pearl].
Conflict of interest statement. None declared.

REFERENCES

1. Caldecott,K.W. (2008) Single-strand break repair and genetic disease. *Nat. Rev. Genet.*, **9**, 619–631.
2. Caldecott,K.W. (2003) XRCC1 and DNA strand break repair. *DNA Repair (Amst.)*, **2**, 955–969.
3. Thompson,L.H. and West,M.G. (2000) XRCC1 keeps DNA from getting stranded. *Mutat. Res.*, **459**, 1–18.
4. Thompson,L.H., Brookman,K.W., Dillehay,L.E., Carrano,A.V., Mazrimas,J.A., Mooney,C.L. and Minkler,J.L. (1982) A CHO-cell strain having hypersensitivity to mutagens, a defect in DNA strand-break repair, and an extraordinary baseline frequency of sister-chromatid exchange. *Mutat. Res.*, **95**, 427–440.
5. Zdzienicka,M.Z., van der Schans,G.P., Natarajan,A.T., Thompson,L.H., Neuteboom,I. and Simons,J.W. (1992) A Chinese hamster ovary cell mutant (EM-C11) with sensitivity to simple alkylating agents and a very high level of sister chromatid exchanges. *Mutagenesis*, **7**, 265–269.
6. Op het Veld,C.W., Jansen,J., Zdzienicka,M.Z., Vrieling,H. and van Zeeland,A.A. (1998) Methyl methanesulfonate-induced hprt mutation spectra in the Chinese hamster cell line CHO9 and its xrcc1-deficient derivative EM-C11. *Mutat. Res.*, **398**, 83–92.
7. Lee,Y., Katyal,S., Li,Y., El-Khamisy,S.F., Russell,H.R., Caldecott,K.W. and McKinnon,P.J. (2009) The genesis of cerebellar interneurons and the prevention of neural DNA damage require XRCC1. *Nat. Neurosci.*, **12**, 973–980.
8. Caldecott,K.W., Aoufouchi,S., Johnson,P. and Shall,S. (1996) XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res.*, **24**, 4387–4394.
9. Pleschke,J.M., Kleczkowska,H.E., Strohm,M. and Althaus,F.R. (2000) Poly (ADP-ribose) binds to specific domains in DNA damage checkpoint proteins. *J. Biol. Chem.*, **275**, 40974–40980.
10. Schreiber,V., Amé,J.-C., Dollé,P., Schultz,I., Rinaldi,B., Fraulob,V., Ménissier-de Murcia,J. and de Murcia,G. (2002) Poly (ADP-ribose) polymerase-2 (PARP-2) is required for efficient base excision DNA repair in association with PARP-1 and XRCC1. *J. Biol. Chem.*, **277**, 23028–23036.
11. El-Khamisy,S.F., Masutani,M., Suzuki,H. and Caldecott,K.W. (2003) A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Res.*, **31**, 5526–5533.
12. Dantzer,F., Amé,J.-C., Schreiber,V., Nakamura,J., Ménissier-de Murcia,J. and de Murcia,G. (2006) Poly (ADP-ribose) polymerase-1 activation during DNA damage and repair. *Methods Enzymol.*, **409**, 493–510.
13. Lan,L., Nakajima,S., Oohata,Y., Takao,M., Okano,S., Masutani,M., Wilson,S.H. and Yasui,A. (2004) In situ analysis of repair processes for oxidative DNA damage in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 13738–13743.
14. Mortusewicz,O., Amé,J.-C., Schreiber,V. and Leonhardt,H. (2007) Feedback-regulated poly (ADP-ribosyl)ation by PARP-1 is required for rapid response to DNA damage in living cells. *Nucleic Acids Res.*, **35**, 7665–7675.
15. Okano,S., Lan,L., Caldecott,K.W., Mori,T. and Yasui,A. (2003) Spatial and temporal cellular responses to single-strand breaks in human cells. *Mol. Cell. Biol.*, **23**, 3974–3981.
16. Li,M., Lu,L.-Y., Yang,C.-Y., Wang,S. and Yu,X. (2013) The FHA and BRCT domains recognize ADP-ribosylation during DNA damage response. *Genes Dev.*, **27**, 1752–1768.
17. Fisher,A.E.O., Hochegger,H., Takeda,S. and Caldecott,K.W. (2007) Poly (ADP-ribose) polymerase 1 accelerates single-strand break repair in concert with poly (ADP-ribose) glycohydrolase. *Mol. Cell. Biol.*, **27**, 5597–5605.
18. Parsons,J.L., Dianova,I.I., Allinson,S.L. and Dianov,G.L. (2005) DNA polymerase beta promotes recruitment of DNA ligase III alpha-XRCC1 to sites of base excision repair. *Biochemistry*, **44**, 10613–10619.
19. Parsons,J.L., Dianova,I.I., Boswell,E., Weinfeld,M. and Dianov,G.L. (2005) End-damage-specific proteins facilitate recruitment or stability of X-ray cross-complementing protein 1 at the sites of DNA single-strand break repair. *FEBS J.*, **272**, 5753–5763.
20. Hanssen-Bauer,A., Solvang-Garten,K., Sundheim,O., Peña-Diaz,J., Andersen,S., Slupphaug,G., Krokan,H.E., Wilson,D.M., Akbari,M. and Otterlei,M. (2011) XRCC1 coordinates disparate responses and multiprotein repair complexes depending on the nature and context of the DNA damage. *Environ. Mol. Mutagen.*, **52**, 623–635.
21. Abdou,I., Poirier,G.G., Hendzel,M.J. and Weinfeld,M. (2014) DNA ligase III acts as a DNA strand break sensor in the cellular orchestration of DNA strand break repair. *Nucleic Acids Res.*, doi:10.1093/nar/gku1307.
22. Woodhouse,B.C., Dianova,I.I., Parsons,J.L. and Dianov,G.L. (2008) Poly (ADP-ribose) polymerase-1 modulates DNA repair capacity and prevents formation of DNA double strand breaks. *DNA Repair (Amst.)*, **7**, 932–940.
23. Caldecott,K.W., Tucker,J.D., Stanker,L.H. and Thompson,L.H. (1995) Characterization of the XRCC1-DNA ligase III complex in vitro and its absence from mutant hamster cells. *Nucleic Acids Res.*, **23**, 4836–4843.
24. Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S., Zacharias,D.A. and Tsien,R.Y. (2002) A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7877–7882.
25. Rulten,S.L., Fisher,A.E.O., Robert,I., Zuma,M.C., Rouleau,M., Ju,L., Poirier,G., Reina-San-Martin,B. and Caldecott,K.W. (2011) PARP-3 and APLF function together to accelerate nonhomologous end-joining. *Mol. Cell*, **41**, 33–45.
26. Breslin,C., Clements,P.M., El-Khamisy,S.F., Petermann,E., Iles,N. and Caldecott,K.W. (2006) Measurement of chromosomal DNA single-strand breaks and replication fork progression rates. *Methods Enzymol.*, **409**, 410–425.
27. Ericsson,U.B., Hallberg,B.M., Detitta,G.T., Dekker,N. and Nordlund,P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.*, **357**, 289–298.
28. Taylor,R.M., Thistlethwaite,A. and Caldecott,K.W. (2002) Central role for the XRCC1 BRCT I domain in mammalian DNA single-strand break repair. *Mol. Cell. Biol.*, **22**, 2556–2563.
29. Marsischky,G.T., Wilson,B.A. and Collier,R.J. (1995) Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins. *J. Biol. Chem.*, **270**, 3247–3254.
30. Rolli,V., O'Farrell,M., Menissier-de Murcia,J. and de Murcia,G. (1997) Random mutagenesis of the poly (ADP-ribose) polymerase catalytic domain reveals amino acids involved in polymer branching. *Biochemistry*, **36**, 12147–12154.

31. Rulten,S.L., Cortes Ledesma,F., Guo,L., Iles,N.J. and Caldecott,K.W. (2008) APLF (C2orf13) is a novel component of poly (ADP-ribose) signaling in mammalian cells. *Mol. Cell. Biol.*, **28**, 4620–4628.
32. Fan,J., Otterlei,M., Wong,H.-K., Tomkinson,A.E. and Wilson,D.M. (2004) XRCC1 co-localizes and physically interacts with PCNA. *Nucleic Acids Res.*, **32**, 2193–2201.
33. Akbari,M., Solvang-Garten,K., Hanssen-Bauer,A., Lieske,N.V., Pettersen,H.S., Pettersen,G.K., Wilson,D.M., Krokan,H.E. and Otterlei,M. (2010) Direct interaction between XRCC1 and UNG2 facilitates rapid repair of uracil in DNA by XRCC1 complexes. *DNA Repair (Amst.)*, **9**, 785–795.
34. Hanssen-Bauer,A., Solvang-Garten,K., Gilljam,K.M., Torseth,K., Wilson,D.M., Akbari,M. and Otterlei,M. (2012) The region of XRCC1 which harbours the three most common nonsynonymous polymorphic variants, is essential for the scaffolding function of XRCC1. *DNA Repair (Amst.)*, doi:10.1016/j.dnarep.2012.01.001.
35. Caldecott,K.W. (2001) Mammalian DNA single-strand break repair: an X-ra(y)ted affair. *Bioessays*, **23**, 447–455.
36. Taylor,R.M., Thistlethwaite,A. and Caldecott,K.W. (2002) Central role for the XRCC1 BRCT I domain in mammalian DNA single-strand break repair. *Mol. Cell. Biol.*, **22**, 2556–2563.
37. Caldecott,K.W. (2014) Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA Repair (Amst.)*, doi:10.1016/j.dnarep.2014.03.021.
38. Ray Chaudhuri,A., Hashimoto,Y., Herrador,R., Neelsen,K.J., Fachinetti,D., Bermejo,R., Cocito,A., Costanzo,V. and Lopes,M. (2012) Topoisomerase I poisoning results in PARP-mediated replication fork reversal. *Nat. Struct. Mol. Biol.*, doi:10.1038/nsmb.2258.
39. Bryant,H.E., Petermann,E., Schultz,N., Jemth,A.-S., Loseva,O., Issaeva,N., Johansson,F., Fernandez,S., McGlynn,P. and Helleday,T. (2009) PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *EMBO J.*, **28**, 2601–2615.
40. Yang,Y.-G., Cortes,U., Patnaik,S., Jasin,M. and Wang,Z.-Q. (2004) Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but compromises the reactivation of stalled replication forks. *Oncogene*, **23**, 3872–3882.
41. Sugimura,K., Takebayashi,S.-I., Taguchi,H., Takeda,S. and Okumura,K. (2008) PARP-1 ensures regulation of replication fork progression by homologous recombination on damaged DNA. *J. Cell Biol.*, **183**, 1203–1212.
42. Le Page,F., Schreiber,V., Dherin,C., de Murcia,G. and Boiteux,S. (2003) Poly (ADP-ribose) polymerase-1 (PARP-1) is required in murine cell lines for base excision repair of oxidative DNA damage in the absence of DNA polymerase beta. *J. Biol. Chem.*, **278**, 18471–18477.
43. Ding,R., Pommier,Y., Kang,V.H. and Smulson,M. (1992) Depletion of poly (ADP-ribose) polymerase by antisense RNA expression results in a delay in DNA strand break rejoining. *J. Biol. Chem.*, **267**, 12804–12812.
44. Vodenicharov,M.D., Sallmann,F.R., Satoh,M.S. and Poirier,G.G. (2000) Base excision repair is efficient in cells lacking poly (ADP-ribose) polymerase 1. *Nucleic Acids Res.*, **28**, 3887–3896.
45. Allinson,S.L., Dianova,I.I. and Dianov,G.L. (2003) Poly (ADP-ribose) polymerase in base excision repair: always engaged, but not essential for DNA damage processing. *Acta Biochim. Pol.*, **50**, 169–179.
46. Ström,C.E., Johansson,F., Uhlén,M., Szigyarto,C.A.-K., Erixon,K. and Helleday,T. (2011) Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate. *Nucleic Acids Res.*, **39**, 3166–3175.